

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

24236

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/68, C07H 21/02, 21/04</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 95/15400</b> <b>(43) International Publication Date:</b> 8 June 1995 (08.06.95)
<b>(21) International Application Number:</b> PCT/US94/13945 <b>(22) International Filing Date:</b> 5 December 1994 (05.12.94) <b>(30) Priority Data:</b> 160,837 3 December 1993 (03.12.93) US <b>(71) Applicant:</b> THE JOHNS HOPKINS UNIVERSITY [US/US]; 720 Rutland Avenue, Baltimore, MD 21205 (US). <b>(72) Inventor:</b> LEVITT, Roy, C.; 1107 Timber Trail Road, Towson, MD 21286 (US). <b>(74) Agents:</b> POSORSKE, Laurence, H. et al.; Banner, Birch, McKie & Beckett, 11th floor, 1001 G Street N.W., Wash- ington, DC 20001 (US).		<b>(81) Designated States:</b> CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> GENOTYPING BY SIMULTANEOUS ANALYSIS OF MULTIPLE MICROSATELLITE LOCI  <b>(57) Abstract</b>  Despite the introduction of molecular methods, such as the polymerase chain reaction, and the discovery of highly polymorphic microsatellite markers, genotyping remains a rate limiting factor in our ability to localize disease genes by linkage. This invention provides a method for genotyping multiple loci by semi-automated fluorescence-based techniques that is both highly accurate and efficient as compared to conventional techniques. The semi-automated techniques developed will be useful in high resolution genomic analyses including: linkage studies, cancer genetics, forensics, and cytogenetics including studies of uniparental disomy or other patterns of chromosomal inheritance.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

**GENOTYPING BY SIMULTANEOUS ANALYSIS  
OF MULTIPLE MICROSATELLITE LOCI**

5

The work leading to this invention was supported in part by Grant No. GM 47145 from the National Institutes of Health. The United States Government may retain certain rights in this invention.

**BACKGROUND OF THE INVENTION**

10 

**Field of the Invention**

This invention is directed to semi-automated methods for linkage mapping of the genome by genotyping of multiple microsatellite loci.

**Summary of Background Information**

15 

For most genetic disorders, there is no known biochemical defect. Consequently, the mutant genes associated with the disease and their disease-causing abnormal gene products are recognized solely by the anomalous phenotype they produce. Identifying the chromosomal localization for the gene(s) that produce these disease phenotypes is often the first crucial step toward isolation and characterization of the mutation(s) by recombinant DNA techniques.

The significance of mapping a gene is perhaps better appreciated when put into context with the human genome project. Consider for a moment that even after every base of the DNA in the entire human genome has been sequenced through the Human Genome Initiative (HGI), and every gene has been localized in this sequence, it may still not be clear which disorder(s) arise from which gene(s). Each disease phenotype will still need to be "mapped" or associated with a particular location in the genome. This is usually carried out by analyzing DNA isolated from blood specimens collected from individuals within families affected by a genetic disorder. Once a disorder or abnormal phenotype has been linked to a particular region on a chromosome, the limited number of genes within this area will permit us to suggest a candidate gene that can contribute to the phenotype. Thus, once the localization of a major disease phenotype to a chromosomal region is confirmed, a few candidate genes can be examined for mutations as well as potential pathogenic mechanisms.

If no genes have been mapped to the region, then linkage studies with closely-spaced surrounding markers can often be used to delineate a large chromosomal interval (1-2 Mb) in which to search for transcribed sequences. This approach (originally termed "reverse genetics") is now generally referred to as "positional cloning". In the past the isolation of candidate genes from these large genomic regions was the rate-limiting step in positional cloning, requiring years of intensive work. However, recent improvements in methods to capture expressed sequences encoded within large genomic segments have been described. Thus, there is now a need for advances in the molecular genetic methods employed in the linkage mapping of disease genes.

### Linkage

The chromosomes are the basic units of inheritance on which genes and DNA markers are organized in a linear fashion (see Figure 1). Linkage is evident when a gene(s) that

- 3 -

produces a phenotypic trait, or a significant portion of the trait, and the surrounding DNA markers are inherited together (cosegregate at meiosis). In contrast, those markers that are not associated with the anomalous phenotype of interest will be randomly distributed among affected family members as a result of the independent assortment of chromosomes and crossing over during meiosis (see Figure 2, compare "A" markers to "B"- "F" markers).

In general, the further a marker, or gene, is from the genetic locus of interest (for example, markers 1 and 4 as compared to markers 2 and 3 in Figure 1), the more likely they will be separated by crossing over at meiosis. The recombinant genotypes produced by crossing over between maternal and paternal chromosomes at meiosis allows us to predict the ordering of genes and markers through the interval under examination. Recombination between the markers 1A and 3A, and 2A and 4A in the affected members in Figure 2, suggest that the mutant gene of interest lies between markers 1 and 4. Thus linkage to a marker of known chromosomal location allows placement of the phenotype on the chromosomal map.

Analysis for testing linkage with use of DNA markers is based on standard likelihood theory. The DNA markers are used to recognize each of the parental chromosomes. Recall that in general each chromosome is inherited independently of any other; and the likelihood of inheriting either chromosome of a pair from each parent is 50:50. Therefore, when a marker is unlinked to the gene(s) producing an anomalous phenotype, one expects both the maternal and paternal chromosomes to be equally distributed in the affected offspring.

Linkage in the human is established by the method of likelihood ratios (see Ott, 1992 "Analysis of Human Genetic Linkage," The Johns Hopkins University Press, Baltimore, for a review). One compares the probability that observed family data, such as that in Figure 2, would arise under one hypothesis (for instance, linkage with no recombination with marker 2

or 3) to the probability that it would arise under an alternative hypothesis (typically, nonlinkage). The ratio of these probabilities is called the odds ratio for one hypothesis relative to the other. By convention, mammalian geneticists prefer the log of the odds ratio, or the lod score. Generally, linkage is considered proven when the odds in favor of linkage versus nonlinkage become overwhelming, or reach 1000:1 (LOD = 3) (see Morton, 1955, *Am. J. Hum. Genet.*, 7:277-318). Linkage is rejected when the odds drop to 100:1 against this hypothesis (LOD = -2). The maximum likelihood estimate is the recombination fraction where the likelihood ratio is largest. Lod scores from multiple pedigrees are thus added until the score grows to 3 (signifying 1000:1 odds) or falls to -2 (indicating 1:100 odds). Linkage can be easily evaluated using likelihood ratios, even in complicated pedigrees, by testing on the computer for these competing hypothesis. Recently, additional strategies have been devised that can handle genetic heterogeneity more effectively (Oh, 1974, *Am. J. Hum. Genet.*, 26:588-597) as well as disorders caused by multiple genes (Lander, et al., 1986, *Proc. Natl. Acad. Sci. USA*, 83:7353-7357).

#### Genotyping With Molecular Genetic Methods

The descriptions of many types of DNA sequence polymorphisms have provided the fundamental basis for our understanding of the structure of the mammalian genome (CEPH consortium map, 1992, *Science*, 259:67-86; Weissenbach et al., 1992, *Nature*, 359:794). The construction of extensive framework linkage maps has been greatly facilitated by the use of these DNA polymorphisms, and has provided a practical means for the localization of disease genes by linkage. The process of linkage mapping in Mendelian and complex disorders using these techniques has been further facilitated by the recent description of a detailed "second-generation" linkage map of the human genome (Weissenbach et al., 1992). In particular the recent description of highly polymorphic PCR-based microsatellite markers for genotyping has greatly

advanced the construction of high resolution linkage maps (Weber and May, 1989, *Am. J. Hum. Genet.*, 44:388-396; Litt and Luty, 1989, *Am. J. Hum. Genet.*, 44:397-401).

The microsatellite markers are highly polymorphic, simple sequence repeat (SSR) markers, generally defined as repeats of 6 bp or less running in tandem for up to 100 bp long (Beckmann, et al., 1992, *Genomics*, 12:627-631). These repeat sequences are flanked by unique DNA sequences that may be identified for each marker location. With primers that correspond to the unique DNA sequence surrounding each marker, the polymerase chain reaction (PCR, see, e.g., Saiki, et al., 1988, *Science*, 239:489) can be used to detect each polymorphism.

This type of genetic marker is abundant and found throughout the genome. SSR may be as frequent as one every 6 kb (Beckmann, et al., 1992). Where SSR markers show considerable polymorphism (differences in the number of repeats) between individuals, the markers can be particularly informative. Many such SSR markers have been isolated throughout the genome, and are well mapped (Weissenbach, et al., 1992). Many of these SSR markers are now available commercially for linkage studies (e.g., from Research Genetics, Huntsville, AL). Those markers which frequently allow the investigator to identify each parental chromosome as unique and to identify each crossover rapidly (see Figure 2) approach the ideal for linkage studies.

Most SSR are (GT)<sub>n</sub> dinucleotide repeat length polymorphisms (see Figure 3). It is estimated that there are about 100,000 of the (GT)<sub>n</sub> type SSR, or one approximately every 30 kb (Beckmann, et al, 1992). Over 1,000 SSR markers have been described to date in the Genome Data Base, October 19, 1993, The Johns Hopkins University, Baltimore, Maryland, and thousands of additional markers are now in development.

It is now well accepted that methods based on the polymerase chain reaction (PCR) and highly polymorphic simple sequence repeat (SSR) markers (e.g. Figure 3) are the techniques of choice for genotyping in linkage studies (Weber, et al., 1989; Litt, et al., 1989; Edwards, et al. 1991, *Am J. Hum. Genet.*, 49:746-56). PCR-based methods are faster and therefore less costly than restriction fragment length polymorphism (RFLP) methods; moreover, they do not require nucleic acid probes, and are more informative in linkage studies. Efforts are underway to develop automated techniques for genotyping that will further improve the efficiency of linkage studies utilizing this type of microsatellite markers polymorphism. The advantages of analyzing multiple polymorphic loci using an automated DNA sequencer were first described by Skolnick and Wallace in 1988 (*Genomics*, 2:273-279). Building on techniques reported by Connell, et al. (1987, *Biotechniques*, 5:342-348), Ziegle et al., (1992 *Genomics*, 14:1026-1031), extended this approach to incorporate automated DNA sizing technology for genotyping microsatellite loci using four color fluorescence-based techniques.

However, the analysis of microsatellite markers still relies on gel electrophoresis which has limited sample handling capacity. Furthermore, the gel electrophoresis of DNA fragments is complicated by problems with gel distortion, such as band shifting that warrant internal size standards and bandmatching software (Lander, 1991, *Am .J. Hum. Genet.*, 48:819-823). Crosstalk or interference during analysis between multiple dyes with spectral overlap is another potential problem when multiple PCR fragments of the same size are to be identified within the same gel lane. Since the processing of gels and the scoring of autoradiographs remains the rate-limiting step in genotyping, methods are being sought that improve the efficiency of sample handling while minimizing errors in data transcription and analysis.



The challenge of mapping the major genes in complex disorders requires efficient and highly accurate methods of genotyping. Recent technological enhancements in molecular genetics have significantly improved our ability to locate disease genes by linkage analysis. However, despite the introduction of molecular methods, such as PCR, and the discovery of highly polymorphic SSR, genotyping is still rate-limiting for localizing disease genes by linkage. The present methods remain highly technical, time-consuming, and expensive.

#### SUMMARY OF THE INVENTION

It is an object of this invention to provide a robust semi-automated protocol for genotyping using multiplex analysis of many microsatellite loci while maintaining, or improving, typing accuracy as compared to traditional methods. It is also an object of this invention to provide a collection of highly reproducible microsatellite markers at approximately 10-50 cM intervals throughout the human genome which can be detectably-labelled.

It is a further object to provide protocols for the reliable use of these marker systems in automated genotyping.

To meet these and other objects, and to better exploit the inherent advantages of fluorescence-based genotyping techniques, this invention provides highly informative SSR markers, assembled into "SETS" that do not overlap in size when separated electrophoretically on an acrylamide gel and that can be labelled with different fluorophores. Each SET contains 6 or more pairs of primers that provide for amplification of markers (preferably 7-8 pairs of primers) that have been labelled with the same fluorophore having a distinct color, separate SETs having different fluorophore labels (e.g., blue, green, or yellow). PCR products corresponding to these SETS are combined into a GROUP for electrophoretic analysis in a single lane. Using this methodology, a GROUP of 18 or more, preferably 21 to 24 dinucleotide

markers can be electrophoresed along with an internal size standard and analyzed simultaneously (multiplexing) in real-time for each individual studied.

In particular, the invention provides a kit for use in automated genotyping within a population comprising four or more GROUPS, each GROUP containing at least three SETS, and each SET in turn comprising at least 6 labelled pairs of primers for amplification of DNA by polymerase chain reaction (PCR), the sequence of each primer pair corresponding to a portion of the unique genomic sequence of a microsatellite sequence (which is made up of a nucleotide repeat sequence flanked by unique sequences), the nucleotide repeat sequence being polymorphic within the population. Amplification of DNA from a human sample by the polymerase chain reaction (PCR) primed with a particular primer pair amplifies the nucleotide repeat sequence and at least some of the immediately adjacent unique sequences of the microsatellite sequence to produce a PCR product identified with the primer pair. The distance in the genome between the microsatellite sequence amplified by one primer pair of the kit and the nearest other microsatellite sequence amplified by another primer pair of the kit is at least 2 centimorgans (cM) and no more than 50 cM. Each SET consists of at least 6 of the primer pairs, where the length of the segment amplified by a particular primer pair (its PCR product) differs from the length of PCR products from all other primer pairs in the SET by at least 5 nucleotides for tetranucleotide repeats, at least 6 nucleotides for trinucleotide repeats and at least 9 nucleotides for dinucleotide repeats. At least one primer of each primer pair is labelled with a fluorescent label that is the same for all primer pairs in the SET. Each GROUP consists of at least three SETS of primer pairs labelled with fluorescent labels, and primers from one SET in the GROUP are labelled with a fluorescent label which fluoresces at a wavelength which is substantially

different from the wavelength at which the fluorescent labels on the primers in each of the other SETS in the GROUP fluoresce.

Where the primers in a single kit cover the entire genome with markers spaced approximately 10 cM apart in the genome, the kit will usually contain at least about 10 GROUPS. In another embodiment, a kit is provided for screening of the genome with individual markers spaced in the genome about 50 cM from the nearest other marker in the kit, and the kit contains at least 4 GROUPS. The invention also provides kits containing fewer GROUPS with primers whose PCR products identify microsatellite sequences found in the genome spaced closely about the locations picked out by screening studies performed using the screening kit.

The invention also provides a method of analyzing genomic DNA for the presence of polymorphisms comprising: extracting DNA from a human sample; combining, in a polymerase chain reaction (PCR) vessel, an aliquot of the extracted DNA, at least one primer pair selected from one of the GROUPS described above, and PCR amplification enzymes; cycling the temperature of each PCR vessel to produce PCR products that can be identified with the primer pair whose sequence corresponds to unique sequence in the amplified DNA, using an annealing temperature at which non-specific annealing is minimized; then combining all PCR products from all PCR vessels containing primer pairs from a single GROUP into a mixture, and subsequently separating the mixture of PCR products electrophoretically by size; and detecting separated PCR products by fluorescence detection at wavelengths corresponding to the fluorescent wavelength for each of the fluorescent labels in the kit. In a preferred embodiment, one primer of each primer pair is labelled with a fluorescent label and the other primer in the pair is labelled with biotin, and a mixture containing all PCR products corresponding to the

- 10 -

primer pairs from a single GROUP is prepared by binding the PCR products to a plurality of paramagnetic beads carrying on their surface a protein which specifically binds biotin (the beads being added to each PCR vessel after amplification), separating the magnetic beads from the PCR reaction medium, then separating the two strands of the amplified DNA segments and combining the strands labelled with a fluorescent label for all primer pairs from one GROUP into the mixture.

The invention also provides a method for selecting a SET of PCR primers for use in automated genotyping comprising selecting at least 6 microsatellite sequences, which contain dinucleotide, trinucleotide or tetranucleotide repeat sequences that are flanked by unique sequences in the human genome, and are polymorphic within the population, the microsatellite sequences being separated from each other by at least 2 centimorgans in the genome, and for each microsatellite sequence constructing primer pairs having the sequence of the unique sequences flanking the microsatellite sequences, so that the primer pairs will direct PCR amplification of DNA segments corresponding to each microsatellite sequence and the length of all polymorphs of the microsatellite sequence amplified by a particular primer pair is detectably different from the length of all polymorphs of other microsatellite sequences amplified by other primer pairs in the SET. The invention also provides a kit for use in automated genotyping comprising at least 10 GROUPS of at least 3 SETS of PCR primers obtained by this method, and a method of analyzing genomic DNA for the presence of polymorphisms comprising amplifying DNA extracted from a human sample using PCR directed by these primer pairs to produce PCR products labelled with detectable labels that are the same for all PCR products from a single SET, followed by separating electrophoretically a mixture containing all PCR products amplified

from the DNA sample by any primer pair of said SET and characterizing the detectably labelled PCR products by length.

The invention also provides a diagnostic method for detection by polymerase chain reaction of genomic rearrangement (including deletions, additions, crossovers and gene amplification), of a genomic region containing at least 6 known loci at which genetic rearrangement is diagnostic for a disease, using a kit comprising at least one SET containing at least 6 PCR primer pairs, the sequences of each primer pair corresponding to the unique sequences flanking one of the loci of genomic rearrangement. The primer pairs in the SET are constructed so that the PCR product amplified by a particular pair of primers corresponds to a DNA segment surrounding one locus of rearrangement with length that is characteristic of a specific rearrangement, and the length of the PCR products amplified by a particular pair of primers differs from the length of all other PCR products amplified by other primers in the SET. DNA from a sample is amplified in a PCR vessel using the polymerase chain reaction (PCR) primed with at least one of the primer pairs of the SET by cycling the temperature of the vessels with an annealing temperature that minimizes non-specific annealing to produce detectably labelled PCR products, and the PCR products for all primer pairs in the SET are detectably labelled with the same label. Labelled PCR products are separated electrophoretically by size from a mixture containing all PCR products amplified from the DNA sample by any primer pair of the SET, and the separated, detectably labelled PCR products are characterized by length. In a preferred mode, all primers in the SET have annealing temperatures within a 4C range, and amplification for all primers in the SET is carried out simultaneously in the same vessel.

The inventor has created a kit comprising SETS of highly polymorphic fluorescent primers specific for microsatellite markers that cover the genome at approximately 10 cM

- 12 -

intervals for linkage studies. A fluorescence-based protocol based on these SETS has been developed for detection of multiple microsatellite markers, and the protocol is accurate as compared to a conventional radiolabeling method that depends on a known DNA sequence ladder and conventional autoradiography for detection. It has now been demonstrated that genotyping by semi-automated fluorescence-based techniques is both highly accurate and efficient. We routinely type 24 fluorescent markers simultaneously using these techniques in my laboratory. The combined analysis of 24 dinucleotide markers in a single gel maximizes the use of automated analysis equipment, such as the Applied Biosystems 373A hardware, by producing PCR products sufficiently small to run the instrument at least twice daily. The methods provided herein may improve productivity by more than an order of magnitude and can be easily adopted to most linkage studies.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the genetic map of the chromosomal region surrounding a putative GENETIC locus. In this example the greater the spacing between markers the more likely recombination will occur during meiosis.

Figure 2 shows segregation data from a fabricated three generation family affected with a genetic disorder for the four markers illustrated in Figure 1. Squares indicate males, circles indicate females. Affected and unaffected family members are indicated by solid and open symbols, respectively. Crossovers that have occurred during meiosis are indicated by the arrowheads. Recombination with markers 1 and 4 from chromosome A exclude a localization for the gene causing this disorder in the region immediately above marker 1 and below marker 4. The region from chromosome A between markers 1 and 4 (including markers 2 and 3) co-segregates with the abnormal phenotype in all the affected individuals in this family but is not

- 13 -

found in any unaffected individuals. These data confirm a localization for the GENETIC locus under study to this chromosomal region.

Chromosomal region 4 of chromosome B from affected individual I-1 occurs in both affected and unaffected offspring in generation II, showing no linkage. The markers used in this demonstration approach the ideal by providing maximal genetic information for every individual studied.

Figure 3 illustrates the most common form of simple sequence repeat. In this individual the marker is heterozygous, or differs in the number of dinucleotides between the maternal and paternal chromosomes. These PCR products would differ in length by 8 nucleotides, and are each easily detected using gel electrophoresis. The solid bars indicate surrounding sequence that is unique (occurs only once in the human genome) and can be used to design PCR primers for amplifying this simple sequence repeat.

Figure 4 shows a cartoon of GROUP 1 markers. Each simple sequence repeat marker is identified on the left, and the size range for known alleles are noted on the right. Each marker covers a region of a chromosome to be examined for linkage with a genetic disorder. The colored boxes refer to the region on the gel where alleles for each marker may be found. The markers are chosen to avoid overlap between these regions. For increased efficiency each SET is labelled with one of three fluorophores -- yellow: tetramethyl-6-carboxy-rhodamine (TMR), blue: 5-carboxy-fluorescein (FAM), and green: 2',7'-dimethoxy-4',5'-dichloro-6-carboxy-fluorescein (JOE); (red 6-carboxy-rhodamine (ROX) is reserved for internal size standards), Applied Biosystems. The products of the PCR amplifications are pooled and subjected to the electrophoresis together. Marker data are derived from the Genome Data Base (GDB), The Johns Hopkins University, Baltimore, Maryland.

Figure 5 shows a typical set of electrophoretograms for GROUP 2 using DNA from a single individual.

Figure 6 shows an electrophoretogram of SET A, GROUP 1 markers from one individual. The size (nucleotides) of each PCR product is given on the X-axis above the electrophoretogram.

Figure 7 A-M provides a listing of the markers in 13 GROUPS each containing 16-24 markers divided into three SETS. The first column gives a locus designation for the marker to identify the entry in the Genbank Data Base which provides the unique sequences surrounding the markers. The unique sequence information can be used to design primers that will direct PCR amplification of the marker. After the locus designation, the size range of the published alleles (in base pairs), the degree of heterozygosity in the population and the chromosomal location are listed, in that order, for each marker followed by the nucleotide sequences of preferred primer pairs, along with their annealing temperatures and preferred choice for labelled primer.

Figure 8 demonstrates the difference in autoradiographic image produced depending on whether the forward or reverse primer is labelled.

Figure 9 shows an autoradiograph of PCR-amplified DNA using the primers of GROUP 2, SET B. The variation in intensity in products of this SET is typical of this type of marker.

Figure 10 shows the effect of varying the amount of paramagnetic beads in a magnetic bead-based recovery from PCR.



**DETAILED DESCRIPTION OF THE INVENTION**

Methods for sequencing DNA, for synthesizing oligodeoxynucleotides of defined sequence, and for separating nucleic acid segments by molecular weight using, e.g., electrophoresis are well known to those skilled in the art and well described in the literature, in, for example, "Molecular Cloning: A Laboratory Manual," Sambrook, et al., eds., Cold Spring Harbor Laboratory Press, 1989. General methods of analyzing DNA by the polymerase chain reaction (PCR) including isolation and preparation of DNA templates, synthesis and labelling of primers, amplification, and analysis of PCR products are also well known and described in the literature, for example in Sambrook, et al., 1989, or in "PCR Protocols: A Guide to Methods and Applications," Innis, et al., eds., Academic Press, 1990. The skilled worker in this art is familiar with these and other methods of manipulating and analyzing DNA, and routine application of such methods within the skill of the ordinary skilled worker is assumed in the following description.

**Semi-Automated Genotyping:**

Despite the improvements in linkage techniques introduced by PCR and SSRs, genotyping remains highly technical, time consuming, and expensive. The application of fluorescence-based technology is one way to further reduce the cost and increase the efficiency of this type of project. Fluorescent labeling of PCR-based markers provides many potential advantages over radio-labels (e.g.,  $^{32}\text{P}$ ) and other labels in common use for PCR markers. Fluorescent labels are nontoxic, stable, and can be combined and analyzed together in a single electrophoretic lane (multiplexing) to provide a many-fold increase in efficiency over standard methods of detection. Fluorescence signals are linear over a much greater range of intensity than conventional autoradiography and other methods of detection in use, providing a better means of

distinguishing between alleles and artifact. Band intensity provides an objective method for distinguishing between alleles and artifacts and may also provide a better means for identifying the products of microsatellite markers that frequently vary significantly in intensity.

Ultimately, real-time fluorescence detection methods may provide a substantial increase in efficiency over standard methods of detection based on radiolabeling. A much larger range of product sizes can be resolved on each gel run as compared to radiolabeling techniques because with the automated, real-time equipment such as the Applied Biosystems Inc., the PCR products pass by the detector toward the bottom of the gel where the band resolution is greatest. Efficiency is further improved by the potential real-time semi-automated detection of alleles. In addition, internal size standards are easily incorporated for reproducibility and the accurate sizing of alleles, avoiding day to day variability. Computerized data acquisition and handling further aid productivity and reduce errors in data entry and manipulation. Ultimately, automation is likely to occur more rapidly with fluorescence-based techniques than with other methods of labeling and detection.

As an initial test of the fluorescence technology, a study was conducted comparing the accuracy and reliability of these methods with  $^{32}\text{P}$  end-labeling (see Example 1). Three markers were chosen because they produce PCR products of the same size range. Products of PCR reactions run with primers complementary to the unique sequences on either side of the SSR for these markers were obtained using primer pairs in which one primer of each pair was conjugated to a fluorescent label. These PCR products were electrophoresed simultaneously in a single electrophoretic lane to test if these genotypes could be accurately determined. Similar to the report by Ziegel, et al., 1992, there was no difficulty in discerning PCR fragments of the same size labelled with different fluorophores.

Determining the size of DNA fragments accurately is critical to genotyping in a number of applications. When parental alleles are available, a simple comparison can determine which, if either, parental allele has been passed on to a child. However, frequently in linkage studies the parental alleles are not available for comparison, and paternity must be questioned. This is also true in DNA forensics, where an unknown must be compared with many others and its size determined unambiguously. The analysis of PCR products that differ grossly in concentration is complicated by bandshifting and other gel related artifacts. The accuracy of this typing procedure must be based on empiric studies of reproducibility using "known" samples as standards. Non-polymorphic internal size standards can be used to remedy these problems (Lander, 1991).

Example 1 demonstrates the accuracy of sizing microsatellite PCR products using a fluorescence-based approach as compared to a conventional radiation-based method using a known sequence ladder. DNA templates may be obtained from the collection of Centre d'Etude du Polymorphisme Humaine, Paris (CEPH) for use as a standard set of alleles to compare these techniques, because there is little question of the genetic identity of each of the individuals in this collection. To avoid ambiguity in genotyping with the fluorescent method, fractional size estimates should preferably be accurate to within 0.5 nucleotides. Variation greater than this could lead to confusion during band matching, after rounding up or down for size estimates provided as a fraction of a nucleotide. Since our analysis suggests that the maximum variation is likely to be less than 0.5 nucleotides (and generally significantly less), the method will be useful in the intended applications.

As shown in Example 1, no sizing errors occurred with the use of the multi-color fluorescence-based technique, showing that this methodology is highly accurate and reproducible

for scoring microsatellite markers. Since the only sizing error resulted from the use of the conventional radiolabeling technique, the fluorescence-based protocol appears at least as accurate as the conventional method. Therefore, this approach appears to adequately compensate for gel distortion and dye related artifacts as compared to radiation labeling techniques.

5 Accordingly, the advantages demonstrated for fluorescence-based techniques may be exploited by the method of this invention, which uses at least 6 highly informative SSR markers assembled into a ladder which we have designated a "SET". Each SSR marker is characterized by PCR primer pairs which have the same sequence as a portion of the unique DNA sequence on the 5' side of the sense and antisense strands, respectively, encoding the repeat sequence at  
10 a particular point in the genome. When the genetic material of a particular individual is amplified by PCR using one of these primer pairs, a segment of DNA corresponding to the sequence of the particular SSR and its unique flanking sequences is produced (the PCR product). The size of the PCR product is dependent both on how much of the unique sequences are covered by the primers in the pair and on the number of times the repeat sequence is repeated.  
15 The number of repeats of the simple sequence at a particular locus varies between individuals (polymorphism), and this polymorphism results in PCR products of varying size for different individuals. Thus the size of the PCR product can be used to determine if two individuals have an allele in common at the genetic locus of the SSR marker.

The spacing in the gel between PCR products identified with different markers is critical.  
20 By carefully selecting the length of the primer sequences for each marker, the PCR products corresponding to each marker in a SET are spaced a critical distance from surrounding markers such that none of the PCR products for the largest known alleles of one marker overlap in size with PCR products for the shortest known alleles of another marker in the SET when separated

on a 6% denaturing acrylamide gel. An additional safety margin should be provided, because rare undocumented alleles (larger or smaller) may occur for any given marker. Size spacing of less than 9 nucleotides between dinucleotide SSR markers increases the likelihood for overlap because 2-4 stuttering bands (each 2 nucleotides apart) below the smallest allele of one marker may overlap with the largest allele of the marker below it. PCR products for trinucleotide repeat sequences and tetranucleotide repeat sequences are not observed to exhibit stuttering bands, so the minimum separation distance above and below the largest and smallest known alleles can be less for tri- and tetranucleotide repeats. Usually, PCR products for trinucleotide repeats in a SET will differ by at least 5 base pairs, and for tetranucleotide markers by at least 6 base pairs. Preferably a SET will contain 7-9 SSR markers, most preferably 8-9 markers. The upper limit on the number of markers in a SET is dependent on the length of the electrophoretic separation.

The PCR product of each primer pair in the SET is tagged with the same label, preferably a fluorescent dye. Usually a fluorescent label is covalently attached to one of the primers in a primer pair. Alternatively, the PCR product may be uniformly labelled by adding one or more fluorescently-labelled nucleoside triphosphates to the PCR reaction. Labelling of the primers may be accomplished by including a fluorescently-labelled nucleotide during synthesis of the primer or by linking a fluorescent label to the primer after synthesis. Fluorophore labels for attachment to nucleic acids, including PCR primers, are readily available in the art. (See, e.g., Nagaoka, et al., (1992) *Chem. Pharm. Bull.*, 40:2559-2561; Giusti, et al., (1993) *PCR Methods Appl.*, 2:223-227; Alexandrova, et al., *Nucleic Acids Symp. Ser. 1991*, p. 277; Schubert, et al., (1992) *DNA Seq.*, 7:273-279; Vu, et al., (1990) *Tetrahedron Lett.*, 31:7269-7272.) Usually the labels contain coupling groups that react with modified nucleotides

- 20 -

of the PCR primers to form covalent links. Attaching such fluorophores to the primers in the SETS of this invention is easily within the skill of the ordinary worker. See, e.g., Levenson and Chang, 1990, "Nonisotopically Labelled Probes and Primers," in PCR Protocols, Innis, et al., eds., Academic Press, NY. Fluorescent labels with non-overlapping emission spectra are also available commercially, for example, from Applied BioSystems, Inc., including 5-carboxy-fluorescein (FAM-blue), 2',7'dimethoxy-4',5'-dichloro-6-carboxy-fluorescein (JOE-green), N,N,N',N'-tetramethyl-6-carboxy-rhodamine(TMR-yellow), and 6-carboxy-X-rhodamine(ROX-red); from Biological Detection Systems, Inc., Pittsburgh, PA (BDS) including nucleoside triphosphates coupled to cyanine dyes that fluoresce in the green or orange region, or Boehringer Mannheim Corporation Biochemical Products, Indianapolis, IN, including fluorescein-5(6)-carboxamidocaproxy-dUTP (yellow), 7-hydroxy-coumarin-3-carboxyl-dUTP (blue), and tetramethylrhodamine-5(6)-amino-thiono-dUTP (red).

Additional suggestions for selecting labels with non-overlapping fluorescent spectra and derivitizing oligonucleotides, with them can be found in Smith, et al. 1986, *Nature*, 321:674-679, incorporated herein by reference. Alternatively, primers (or PCR products) may be labelled with biotin (see, e.g., Innis, et al., "PCR Protocols," Academic Press, NY, 1990, pp. 100-103) and then streptavidin coupled to a particular fluorescent dye added to all of the PCR products of a particular SET. Variations of these labelling methods or similar methods known to those skilled in the art may be used, so long as all PCR product for markers in one SET are labelled with the same label.

SETS, each labelled with a different fluorophore, can be pooled into a collection of markers that we have termed a "GROUP." The number of SETS in a GROUP will depend on the availability of distinct labels. PCR products for each SET in the GROUP will usually be

labelled with fluorophores that emit light at a wavelength substantially different from the wavelengths emitted by fluorophore labels of the other SETS in the GROUP, where "substantially different" means sufficiently distinct to be distinguished by the detection means chosen for detecting PCR products after electrophoresis. For example, three commercially available fluorophores, referred to as TMR, FAM, and JOE (Applied Biosystems), have different colors which are yellow, blue, and green, respectively.

Using this approach we have analyzed as many as 24 SSR markers in a single electrophoretic lane using three distinct fluorescent labels to label three SETS in the GROUP (see e.g. Fig. 4). In a preferred mode, these fluorescent PCR products may be separated on an automated electrophoresis systems, such as the Applied Biosystems 373 sequencer with internal size standards in each lane (labelled, for example, with ROX (red dye), Applied Biosystems) and analyzed using, e.g., GeneScan 672 software (Applied Biosystems) (Ziegle, et al., 1991, Miami Short Rep., 1:70) and scored using GENOTYPER software (Applied Biosystems), with data displayed as an electrophoretogram or in a spread sheet format. Gel band fluorescent intensities and peak areas provide an objective method of distinguishing alleles from artifact (stuttering bands). A typical electrophoretogram from a single individual for SET A GROUP 1 is illustrated in Figure 6.

#### Marker Selection and Development:

The human genome is estimated to be approximately 3000 cM in length. Therefore, to adequately "cover" the entire genome at 10 cM intervals will require approximately 300 highly informative well spaced markers. An alternative estimate obtained by summing the meiotic maps from all the chromosomes suggests that the genome is approximately 5000 cM in length (NIH/CEPH Collaborative Mapping Group, 1992, *Science*, 259:67-86). Adequate "coverage"

of the entire genome based on this size estimate at 15 cM intervals (which would allow testing for linkage without using a prohibitively large number of families) will require about 333 highly informative well spaced markers.

Characteristics of preferred markers can be summarized as follows: unique sequence  
5 surrounding the marker is available for use in designing primers, they have been sized accurately, the heterozygosity value is known, and each marker has been carefully localized. Over 1000 SSR markers, including the surrounding unique sequence and chromosomal location, have been described to date in the Genome Data Base (GDB), October 19, 1993, The Johns Hopkins University, Baltimore, Maryland. In contrast to older approaches, such as RFLP, many  
10 of the preferred SSR markers are heterozygous (alleles differ at a particular locus) > 50% of the time and therefore are highly informative for linkage studies. Each allele of the markers used in the method of this invention will be easily detectable after amplification by PCR as a predictable component of a complex image or signature by 5' end labeling with  $^{32}\text{P}$ , labeling with fluorescence, or by a variety of other methods. Most preferably, the markers also produce  
15 an easily scored product or simple pattern of stutter bands that are the signature of mononucleotide and dinucleotide repeats.

Most dinucleotide repeats produce two or three smaller less intense products or "stutter bands" (Weber, 1989). These are artifacts produced during PCR, and are less common in PCR of tri- and tetranucleotide repeats. Although these stutter bands have been generally considered  
20 undesirable, they can be quite helpful to the investigator (or computer) during the scoring of genotypes by allowing for the identification of 'false' bands (background bands due to non-specific annealing). Each allele can then be easily scored by 5' end labeling with  $^{32}\text{P}$  or fluorescence after amplification by PCR, as a predictable component of a complex image.



Background bands are generally not associated with stuttering artifacts. Because artifacts due to nonspecific annealing are difficult to eliminate entirely from a PCR reaction, the adaptation of a similar protocol for the multiplex semi-automated genotyping of tri-, and tetranucleotide repeats may be more problematic. The method of this invention reduces artifacts due to non-specific annealing by control of the annealing temperature for respective primers during temperature cycling.

The use of dinucleotide SSR is preferred in the method of this invention, because the potential advantages for automated genotyping may not be so easily incorporated into practice for mono-, tri- and tetranucleotide repeats. PCR products of trinucleotide and tetranucleotide repeats lack the unique "stuttering" signature of dinucleotide repeats, making it difficult for the computer to distinguish real alleles from artifacts produced by nonspecific annealing during PCR. Although a simple set of PCR products are produced as alleles (little or no stuttering) from tri- or tetranucleotide SSRs, it is often difficult to eliminate other PCR artifacts completely. These PCR artifacts are not easily distinguished from "false" bands when large numbers of PCR products that vary significantly in intensity are combined as described by this method. The unique signature derived from the stuttering bands of dinucleotide repeats provides a simple means of distinguishing real products (alleles) from artifactual bands.

Furthermore, the cost of the hardware is generally considered the limiting factor when adopting the fluorescent approach. Tri- and tetranucleotide markers generally require a significantly larger fraction of each gel because alleles span a much larger size range. Thus longer run time is required, and fewer markers can be resolved per gel. The cost of the hardware becomes readily affordable if one considers the utility and throughput of such an instrument when used according to the method of this invention. However, the use of fewer

markers per lane (i.e., tetranucleotide repeats) would substantially reduce the cost effectiveness of the hardware by reducing efficiency.

Finally, far fewer of tri- and tetranucleotide markers have been fully characterized at present. Thus, the availability of well-characterized primers which can be assembled into SETS and GROUPS remains another limiting factor at present.

#### Construction of Marker SETS:

The selection of markers for inclusion in each SET is based on the need to: maximize heterozygosity values (genetic informativeness), place the marker within a SET based on the size of the PCR products (alleles produced must not overlap with those of the marker above or below it), and the location of the marker in the genetic map (ideally we would have 450-500 markers placed 10 cM or less apart). The PCR products corresponding to markers within a SET are sized to assure that infrequent alleles and stutter bands do not produce overlap between the markers (compare e.g., Figures 4 and 6). PCR products for SETS of dinucleotide markers differ by approximately 9 nucleotides, preferably, at least 10 nucleotides, in length. When necessary, new oligonucleotide primers based on the unique sequence surrounding a polymorphic marker are designed and synthesized to assure that the PCR products do not overlap during electrophoresis.

Figures 7A-M show 289 SSR markers that have been selected and combined into 11 GROUPS of 21-24 markers and 2 incomplete GROUPS of 16 markers so that markers in each GROUP can be separated and analyzed simultaneously. The selected markers cover the genetic map on average once every 10 cM. Most are heterozygous greater than 70% of the time. In a preferred embodiment, each SET is composed of 8 markers from multiple linkage groups (see, e.g., Figure 7B-H). Most preferably, SETS of markers are part of a single linkage group (i.e.

a single chromosome), but this may require significant additional labor because fewer existing primers will be suitable.

Additional or alternative SSR loci to assemble into GROUPS of markers may be found in GDB. Loci listed in GDB can be arranged on the genetic map by using map location  
5 information in GDB. Additional or alternative primers may then be designed using information on the surrounding DNA sequence available in Genbank, based on the locus designations from GDB. GROUP 1 markers (Figure 7A) are currently performing well in multiple laboratories.

In many cases new oligonucleotide primers must be designed from the sequence surrounding each marker to produce PCR products that fit between the products of the markers  
10 above and below it without overlap. The new primers can readily be designed from the known sequence surrounding the SSR. Criteria for selecting a sequence to be synthesized as a PCR primer are well known (see, e.g., Sambrook, et al., and Innis, et al., especially p. 9). Preferably, the unique primer 3' sequence should contain at least 7 nucleotides, the  $\Delta G$  threshold should be at least -1.0 kcal/mol, most preferably -1.4 kcal/mol, and duplex formation  
15 should be avoided, the maximum length of duplex not exceeding 2 base pairs. The sequence of preferred primers will also minimize or eliminate self-complementarity, hairpin formation, and false priming. Once the sequences of candidate primers are chosen, synthesis is readily accomplished by standard methods (see, e.g., Sambrook, et al.).

#### **Optimization of PCR Conditions and Appearance on the Gel:**

20 These new primers must be tested to assure that they produce an easily scored collection of products of the correct size. Scoring may be easier if the label is on one primer rather than the other for particular markers (see, e.g., Figure 8). Primers developed for dinucleotide markers may perform well in the PCR reaction, but produce products unacceptable for

- 26 -

genotyping (single base stuttering bands, stuttering bands of equal intensity with true alleles, or stuttering bands that are larger than the correct allele), and such primers should be avoided.

For best results, the PCR conditions for each marker should be optimized to eliminate any artifactual PCR products due to nonspecific annealing that may complicate the analysis of a GROUP of combined markers. In particular, the temperature of the annealing phase of each PCR cycle should be optimized for each primer pair. Accordingly, the annealing phase temperature is set relatively high, so that specific hybridization occurs, but non-specific hybridization between the template DNA and the primers is minimized. Usually, the selectivity provided by this optimization is preserved in the method of this invention by limiting the number of primer pairs in any PCR reaction vessel to those whose optimized annealing temperature is the same or nearly the same. Preferably, all primer pairs in the same PCR vessel have annealing temperatures within 4C of each other. At one extreme, an entire 96 well plate is dedicated to PCR reactions using primers for a single marker. (When genotyping is preformed for a large number of individuals, using a separate plate for PCR reactions for each marker will not reduce efficiency.) Alternatively, each PCR vessel on a plate has only one primer pair, but the plate contains vessels having different primer pairs, so long as all primer pairs on the same plate have annealing temperatures within 4C. In a preferred mode, all of the primer pairs for a SET or even a GROUP are constructed to have optimized annealing temperatures in a narrow range, most preferably 4°C, and all of the primers are present in a single PCR reaction vessel, obviating the need to mix the individual PCR products prior to electrophoretic separation.

In addition, each marker should be evaluated to assure it is sized correctly within the SET and that the alleles can be easily scored as distinct products. Furthermore, reported heterozygosity values are usually verified using a population of unrelated individuals. The same

DNA templates provided herein may be used as controls for verification of protocols and quality assurance. Preferred controls include CEPH parents (BIOS corporation, New Haven, Conn.; Cell Repository, Camden, N.J.), such as families 1331, 1347, 884, for which reference alleles are known (see, Weber, et al., and Genethon Microsatellite Map Catalog, Genethon Human Genome Research Center, Evry, France). Pooled DNA from volunteers who have donated blood that has been purified as described in the EXAMPLES may be used as well.

This optimization process requires the synthesis of oligonucleotide primers, dilution and aliquoting of primers, identification of the appropriate annealing temperature ( $T^{\circ}$ ) and PCR protocol, electrophoresis of the products, autoradiography and data analysis. If labelled primers are used for detection of products, 5' end labeling of both primers should be tested to determine which one produces the best image<sup>1</sup>. The size of the PCR products from each marker should be verified experimentally to assure that it does not overlap with the products of the surrounding markers in the same SET. As a control for this purpose, PCR products from a pool of DNA samples from a population of unrelated individuals may be electrophoresed against a DNA sequence ladder. In a preferred mode the test pool will contain at least 50 chromosomes.

Initial characterization of primers for each SSR marker may be performed with <sup>32</sup>P labels because this is less costly, but the smooth adaptation of fluorescent-based techniques for genotyping with markers that have been optimized using <sup>32</sup>P is also dependent on assuring the PCR products labelled with a fluorescent dye perform as expected during PCR and analysis. Therefore, the reliability of the developed protocol should be checked by electrophoresis of DNA samples labelled by PCR with the fluorescent labels.

---

<sup>1</sup> Frequently the image produced by labeling one of the pair of primers is blurred, see, e.g., Figure 8.

The PCR products of different microsatellite markers frequently vary significantly in intensity (see, e.g., Figure 9). The sizing of fluorescent PCR products of grossly different concentrations is potentially complicated by sample overloading, causing spectral interference between the dye labels during analysis. There was no interference in the detection of the overlapping products using the four dyes in Examples 1 or 5, because the concentration of each PCR product was determined and adjusted to prevent overloading. However in our experience this can become a problem when working routinely with 21 to 24 pooled markers.

Overloading can lead to artifacts that become especially troublesome when they are interpreted as internal size standards. To prevent the inaccurate sizing of the products by the GeneScan 672 software, we have found that the selection of the standard peaks must be carried out manually. During large scale applications, such as in our linkage studies, this may become a serious problem. Moreover, it is often impractical to estimate the concentration of each of the fluorescent products in order to adjust the concentration of the individual samples to be pooled. Generally adjustments in the volumes for each marker can be made for all the samples by estimating the relative intensity of the marker within a SET. This is easily accomplished by referring to the data table of fluorescent band intensities or by viewing the electrophoretogram directly.

In a preferred mode, PCR products are recovered and combined into a mixture containing the GROUP by a simple protocol that uses magnetic separation technology to purify the fluorescent PCR products and which restricts the total amount of product pooled to prevent overloading. Magnetic separation provides simple separations based on specific binding interactions without the need for expensive centrifuges. Saturation binding to a limited amount of paramagnetic beads can be used to control the amount of labelled PCR product carried

forward in the analysis. Relative intensity may be adjusted by this means and overloading may be avoided.

In a preferred embodiment, one primer is labelled with a component that will bind to magnetic microbeads, for example biotin-labelled primers will bind to streptavidin-coated magnetic beads. Methods for labelling primers with biotin are taught in, e.g., Innis, et al., "PCR Protocols," 1990, pp. 100-103 and references cited therein. Magnetic beads coated with streptavidin are commercially available (Dynabeads™) and procedures for separation are described in, e.g., "Magnetic Separation Techniques Applied to Cellular and Molecular Biology," Kemshead, et al., eds., Wordsmiths' Conference Publications, Somerset, U.K., 1991.

A fixed amount of magnetic beads are added to the PCR reaction after amplification using primers that will bind to the magnetic beads. The magnetic beads with the PCR product attached are separated from the remainder of the PCR reaction mixture, including salts and unused, detectably-labelled primer, and then the PCR product is recovered from the magnetic beads (for example, by separating the strands, leaving one strand attached to the bead and recovering the other strand whose primer carries the detectable label).

Alternatively, the entire PCR product may be labelled by including biotinylated UTP in the PCR reaction medium as described by Dennis, et al., 1990, in "PCR Protocols," Innis, et al., eds. The PCR product can be bound to the beads for purification from the PCR reaction mix and excess primer, and subsequently recovered from the beads by, for example denaturation of streptavidin. In another alternative mode, paramagnetic beads which have attached to their surfaces single stranded DNA corresponding to a part of the sequence of the PCR product may be added to the PCR reaction mix at the end of amplification, followed by cycling above the melting temperature, reannealing and then separating the paramagnetic beads and any other DNA

strands annealed to the beads from the reaction mix. Labelled strands can then be recovered from the beads, as above.

Selection of SETS and GROUPS of fluorescent SSR markers covering the human genome (approximately 300) can be completed in approximately 6-9 months, using the procedures provided herein. Preferably, additional fluorescent markers will be developed (approximately 500 SSR markers) providing a higher resolution tool for gene mapping. The resolution of this marker collection will approach 10 cM and will preferably cover the telomeres which will better assure linkage detection in complex non-Mendelian disorders like asthma and diabetes.

The development of a common index set of fluorescent markers that can be used in multiple laboratories simultaneously should provide certain advantages in genomic studies. Typing these common index loci in a number of different populations afflicted with the same disorder will facilitate the comparison of linkage results and provide the information required for the eventual application of these techniques to forensic medicine.

The method of this invention offers several significant advantages over a similar strategy adopted by Diehl et al., 1991, *Am. J. Hum. Genet.*, 47:177. Spacing markers in a SET according to this invention avoids overlap, providing improved discrimination among markers and between markers and artifacts. As many as eight or more markers may be incorporated into a SET. When necessary, new oligonucleotide primers based on the unique sequence surrounding a polymorphic marker can be designed and synthesized as taught herein to assure that the PCR products do not overlap during electrophoresis. Errors introduced by sample handling may also be minimized by storing DNA from each individual to be studied in a 96-well format. Our protocol preserves the integrity of a 96 well format including PCR amplifications, product pooling, and sample purification, thereby minimizing sample handling and errors introduced by



excessive sample manipulations. In a preferred mode, efficiency is further aided by the transfer of a row of samples by multichannel pipette.

The combined analysis of multiple markers maximizes the use of the Applied Biosystems 373 sequencer or similar automated analysis hardware. Since the capacity of the 373 sequencer is 36 lanes per gel, 864 genotypes (1728 alleles) can be analyzed routinely from one gel using the semi-automated method of this invention. A typical linkage study would include about 100 families or about 500 individuals. For a 5-year study including about 300 markers, approximately 180 gels, or about 3 gels per month, will be required. By using the method of this invention, at least 2 gels per day can be run per 373 sequencer. Thus, up to 12 investigators can be accommodated on one instrument, which substantially reduces the cost per investigator.

The method of this invention can also increase the efficiency of diagnostic studies of the genome, when the desired diagnostic procedures involve the detection of genetic changes that affect the length of genomic DNA at 6 or more locations. Such changes include additions, deletions, intra-and interchromosomal crossover, gene amplification and similar gene rearrangements. The loci of many such rearrangements are known and associated with many diseases, especially cancers and metabolic errors inherited recessively. PCR using primer pairs which direct amplification of a DNA segment including one of these loci can be used diagnostically where the rearrangement associated with the disease causes a change in the length of the PCR product. A SET of primers designed according to the principles above can be used in the production of PCR products that can be analyzed electrophoretically in a single lane, for more efficient use of electrophoresis and analysis equipment.

### EXAMPLES

The following examples describe particular embodiments within the broader invention. These embodiments are described for illustrative purposes only, without intention to limit the invention.

#### 5 EXAMPLE 1

As an initial test of the fluorescence technology, a study was conducted to compare the accuracy and efficiency of these methods with a conventional radiation-based method. Three microsatellite loci producing PCR products that overlap in size were chosen to compare the accuracy of genotyping by fluorescence versus radiolabeling. Discrepancies between the  
10 genotypes derived from each technique were resolved by repetition. To estimate the variation in sizing of the fluorescence-based technique certain samples were loaded on 3 or more gels for comparison. DNA from CEPH (Centre d'Etude du Polymorphisme Humaine, Paris) families 884, 1331, 1332, 1333, 1362 were amplified for Marshfield markers, mfd 1 (176-196bp), mfd 59 (175-195bp), and mfd 154 (186-204bp) using the polymerase chain reaction (PCR).

15 *Fluorescent techniques:* The forward and reverse primers were each labelled at the 5' end for detection by autoradiography with [<sup>32</sup>P] γATP(6000 Ci/μmole) using polynucleotide kinase. A primer was selected from each marker for fluorescent labeling on the basis of the image of the products (see Figure 8). The optimal annealing temperature was selected for each marker empirically by selecting a temperature that eliminated nonspecific annealing or artifactual  
20 (background) PCR products. Fluorescent labels were attached at the 5' end via phosphoramidate derivitization using Aminolink 2 (Applied Biosystems). Primer B (see Figure 10) for mfd 1 was labelled yellow (TMR), primer A (see Figure 10) for mfd 59 was labelled blue (FAM), and primer B (see Figure 10) for mfd 154 was labelled green (JOE). PCR conditions were: 0.4 μM

- 33 -

primers, 1.5  $\mu$ M  $MgCl_2$ , 50  $\mu$ M KCl, 200  $\mu$ M dNTPs and 0.5 units *Taq* polymerase (final concentrations); 94°C for 10 min; followed immediately by 30 cycles of 94°C for 30 sec; 58°C (mfd 59, mfd 154) for 30 sec or 60°C (mfd 1) for 30 sec; and 72°C for 30 sec; followed by 72°C for 7 min. PCR was carried out in a volume of 12.5  $\mu$ l using 25 ng of CEPH DNA.

5 CEPH DNA was stored in a 96 well microtiter plate (Perkin Elmer/Cetus). Amplifications were performed in 96 well microtiter plates using a Perkin Elmer/Cetus Model 9600 thermalcycler and accessories, maintaining the integrity of the 96 well template. Five microliters were combined from each marker for each CEPH individual using a multichannel pipette (Transferpette-8, Brinkman). The pooled PCR products were desalted by adding 2 volumes of  
10 sterile deionized distilled water ( $ddH_2O$ ), ice cold ethanol (100%) equal to the total volume, and chilling for 30 minutes at -70°C. The microtiter plate was spun at 4°C at 1400XG for 2 hours in a Beckman Model GS6R centrifuge. The supernatant was aspirated, the pellet was washed once with 1.5 volumes of ice cold ethanol (70%), and the plate centrifuged 30 minutes at 1400XG at 4°C. The supernatant was aspirated and the plate was air dried. Pellets were  
15 resuspended in a volume of sterile  $ddH_2O$  equal to the starting volume (pool).

Radiolabelled products were separated by conventional electrophoresis and scored manually from autoradiographs. Fluorescent PCR products were separated on a 373 sequencer with internal size standards in each lane (GeneScan 2500-ROX; Applied Biosystem) and analyzed using GeneScan™ 672 software (Applied Biosystems). Each sample (representing 0.5  $\mu$ l of each  
20 product) was heated to 99°C after adding 1  $\mu$ l of the internal lane size standards (GeneScan 2500-ROX, Applied Biosystems) and 2  $\mu$ l formamide/EDTA loading buffer, until the total volume was reduced to 2-3  $\mu$ l. Electrophoresis was carried out using 6% acrylamide (Biorad),

- 34 -

8 M urea (Ultrapure, USB) gels in 1 X TBE. The reduced volume was loaded and run for 4-8 hours on a model 373 Sequencer (Applied Biosystems) using a 24 cm well to read distance.

The size of the PCR product is determined by reference to the internal lane size standards (Carrano et al. 1989, *Genomics*, 4:129-136). The size standard ROX-2500 (Applied Biosystems) including fragments: 37, 94, 109, 116, 172, 186, 222, 233, 238, 269, 286, 361, and 479 nucleotides in length was used with modifications. PCR fragments 61 and 68 nucleotides in length were gel purified, labelled by aminolinking with ROX, and added in equal volumes to the ROX-2500 standards. These fragments were added because desalting by ethanol precipitation recovers the unused PCR primers with the products. The intense peak produced by the unincorporated labelled primer is seen in the standards because of interference between dyes and obscures the detection of the 37 nucleotide standard fragment. Therefore, we have modified the GeneScan-2500 standards to provide a fragment of known size labelled with ROX to accurately estimate the length of the smallest alleles.

The GeneScan 672 (version 1.0) software recognizes any peak labelled with ROX, computes a calibration curve based on a second-order least-squares fit, and uses these data to estimate the allele sizes of the PCR products (Ziegle et al. 1992). Data from each lane can be analyzed independently, or four lanes of data for a single fluorescent dye can be displayed simultaneously to compare individuals within a family. Allele sizes in nucleotide bases, the genotypes, are assigned by interactively distinguishing major peaks from background artifacts. The scale on the display can be adjusted to analyze peaks with differences in fluorescent intensity. The intensity of each fluorescent band and peak areas provide an objective method of distinguishing alleles from artifact (including stuttering bands). Allele sizes can be transferred to a spreadsheet database for linkage or a multicolor electrophoretogram.

SUBSTITUTE SHEET (RULE 26)

mfd 1, mfd 59, and mfd 154 PCR products overlap in size (175-204) bp (see Figure 10).

There was no evidence of interference between the dyes even when there was complete overlap during the electrophoresis of PCR products, similar to that reported by Ziegel et al., 1992. In our experience, interference between dyes does become a problem with overloaded samples.

5 A comparison of the genotyping results of the radioactive and fluorescent labeling methods revealed 4 discrepancies out of 462 possible comparisons (alleles) (see Table 1). One transcription error occurred in the manual data manipulation of the fluorescently labelled products. There was no interference between fluorophores with the detection of the overlapping products using the four dyes. No sizing errors were attributed to the fluorescence-based  
10 technique and each marker displayed Mendelian inheritance. The average size variation across all comparisons was 0.28 nucleotides. However, the maximum difference (range) found for any of the 462 comparisons was 0.47 nucleotides (see Table 2). Generally sizing varied less within a gel than between gels. The variation in the size of the alleles was similar when comparing each of the individual markers. The remaining discrepancies occurred with the use of the  
15 standard radioactive-based protocol and represented an error rate of less than 1%. Inaccurately sized PCR products and sample misloadings produced mistypings with the conventional technique (see Table 1). In general, fluorescent internal size standards provided more precise sizing than did radiolabeling. These data demonstrate both improved accuracy and efficiency for typing SSR markers with use of fluorescence-based techniques.

- 36 -

TABLE 1

CEPH DNA/Marker	Genotype Radiolabelled	Genotype Fluorescence	Explanation
884-18/mfd 1	178,192	178,194*	size estimate error
1331-16/mfd 59	179,179	179,185*	gel loading error
1331-17/mfd 59	179,170	179,185*	gel loading error
61332-15/mfd 154	185,200*	200,200	recording error

\* indicates correct score by length in nucleotide residues

TABLE 2

COMPARISON	RANGE (in nucleotides)		
	Maximum	Average	Standard Deviation
intergel <sup>571</sup>	0.47	0.28	.08
intragel <sup>571</sup>	0.42	0.18	.07
mfd 1 <sup>247</sup>	0.35	0.19	0.1
mfd 59 <sup>177</sup>	0.37	0.15	.08
mfd 154 <sup>147</sup>	0.42	0.23	.06

Superscripts indicate number of samples

**EXAMPLE 2**Mapping with Fluorescent Primers

Genomic DNA is isolated as described by M.J. Johns, et al., *Analytical Biochem.*,  
180:276-278 (1989).

To minimize sample handling, DNA templates can be stored in a 96 well grid (e.g.,  
Perkin Elmer/Cetus). The integrity of the grid may be maintained throughout the protocol to

- 37 -

avoid errors introduced by manual pipetting and sample handling. Multichannel pipetting from a 96-well grid expedites sample handling while minimizing human errors.

PCR is performed in a reaction volume of 12.5  $\mu$ l, containing 50 $\mu$ M dATP, dGTP, dTTP, dCTP; 0.07 $\mu$ M of the labelled oligonucleotide primer, and 4  $\mu$ M of the unlabelled primer. Taq polymerase (Perkin-Elmer\Cetus) 0.5 units is added on ice. PCR will usually be performed in a thermalcycler, e.g., a Perkin-Elmer\Cetus 9600 thermalcycler. Standard thermalcycler settings are 94°C for 10 minutes, followed by 30 cycles 94°C for 30 seconds, 30 seconds at average annealing temperature for the primers and 72°C for 30 seconds; final extension is at 72°C for 7 minutes.

Labelled PCR products are purified by co-precipitation in EtOH. 24 markers may be co-precipitated simultaneously in the 96-well format using ethanol. Ethanol precipitation desalts the products but copurifies the primers. The labelled primer peak produces an enormous signal that complicates the analysis of products under 93 nucleotides in length because it interferes with the 37 nucleotide ROX GeneScan-2500 standard. As an alternative, internal standards may incorporate fragments that are 50, 60, and/or 70 nucleotides in length in addition to the GeneScan 2500 standard fragments or an equivalent set of fragments.

The amplified products are analyzed by denaturing gel electrophoresis (Sambrook, et al.). Loading buffer (2X concentration) is added to an equal volume of the PCR reaction, and the PCR reaction is loaded on a 6% polyacrylamide gel. Radioactive products will be sized against a sequence ladder; the gels are dried and then exposed to Kodak XAR film for 4-24 hours with or without intensifying screens. Fluorescent labelled PCR products may alternatively be analyzed by semi-automated detection using, e.g., an ABI 373A automated sequencers and GeneScan 672 software from Applied Biosystems, Inc.

**EXAMPLE 3**

PCR products are produced as in Example 2 and then purified and combined for electrophoresis using a magnetic bead protocol in place of EtOH precipitation. One of each pair of primers is labelled with biotin and the other with a fluorescent label as above. Double stranded PCR products are purified using streptavidin conjugated to paramagnetic beads to bind the primer 5' labelled with biotin. This procedure may be easily adapted to the 96-well format in any laboratory without expensive centrifuges. After the DNA bound to magnetic beads is separated from the PCR reaction media, the two strands are melted and separated, and the strand labelled with the fluorescent primer is pooled with other labelled strands of its GROUP for electrophoresis. The result of increasing the amount of beads used for separation of a single PCR product from its PCR reaction mix is shown in Figure 12.

**EXAMPLE 4****<sup>32</sup>P OPTIMIZATION OF PRIMER SETS**DNA Templates

CEPH parents and/or unrelated volunteers as controls may be tested. In addition, we usually include one "no DNA" control and one reference individual (alleles known) on each plate. To maximize the use of resources, each marker may be optimized, using 12 wells or less of a 96-well plate. Eight markers are amplified per plate at a single temperature. Alternatively, a thermalcycler with a smaller sample capacity may be used.

The 5' end of the primers to be tested is labelled with <sup>32</sup>P using the polynucleotide kinase reaction. Mix 5  $\mu$  sterile ddH<sub>2</sub>O, 2.8  $\mu$ l 5x kinase buffer (250 mM Tris, 50 mM MgCl<sub>2</sub>, 50 mM DTT, 0.25 mg/ml BSA), 6.0  $\mu$ l 10  $\mu$ M primer, 0.8  $\mu$ l T<sub>4</sub> polynucleotide kinase, and 3.0  $\mu$ l  $\gamma$ <sup>32</sup>P



ATP (6000 Ci/mmol). Incubate at 37° for 1 hour, then add 26 µl sterile ddH<sub>2</sub>O, spin through select D column (Five Prime Three Prime) loaded with P4 Bioigel (BIORAD) according to the manufacturers recommendations. The labelled primers may be stored at -20°C.

For optimization, set up simultaneous PCR reactions as described in Example 2, using  
5 DNA templates described above (e.g., 2 CEPH (1331-1, 1347-02), 1 pooled sample (50 chromosomes), 1 no DNA). Perform PCR at the annealing temperature ( $T^{\circ}$ ) calculated as follows

$T^{\circ} = 2(A+T) + (G+C)$  (If the calculated temperatures for 2 primers differ greatly, for example 54° and 64°, begin closer to lower  $T^{\circ}$ )

10 Check the amplified PCR product for artifact by electrophoresis on 6% gel. Continue optimization of the selected <sup>32</sup>P-labelled primer with control individuals, increasing the annealing temperature in 2° increments until nonspecific products are eliminated. On average, determinations at approximately 4  $T^{\circ}$  values are required to optimize each primer.

When all markers from a SET are optimized (usually 8 markers), 3 µl from a pool of  
15 PCR product of DNA from unrelated individuals using primers for each marker in the SET is combined with an equal volume of loading buffer (2X concentration). Seven µl (or maximum well volume) of the combined mixture is loaded on a gel and electrophoresed. This last check on size and product intensity assures that the markers are robust and are spaced about 10 nucleotides apart. The primer sequences may then be used to synthesize fluorescent/biotinylated  
20 products.

#### EXAMPLE 5

A protocol extending this approach to include up to 24 microsatellite markers in each electrophoretic lane was tested as follows. The selection of markers was based on the need to:

maximize heterozygosity (genetic informativeness), distribute markers across the entire genetic map, and the placement of the marker within a SET based on the known size of the PCR products (alleles and stuttering bands produced must not overlap with those of the marker above or below it).

5           Highly informative microsatellite markers were assembled into a ladder or "SET". Each marker in a SET is spaced a distance of at least 9 nucleotides from surrounding markers such that none of the PCR products overlap in size when separated on a 6% denaturing acrylamide gel. Since many dinucleotide repeats produce a complex pattern of 3 or more stutter bands, this spacing is critical to assure that more intense stutter bands from an upper marker will not be  
10       misinterpreted as a product from a lower marker. In addition, new alleles both larger and smaller than the reported product sizes for this type of marker have occasionally been discovered. Each SET was labelled with one of three different commercially available fluorophores (TMR, FAM, and JOE; Applied Biosystems). The fourth fluorophore (ROX) was reserved for the internal size standard. Three SETS each labelled with a different fluorophore  
15       were pooled into a collection of markers we have termed a "GROUP".

New primers were designed as necessary using OLIGO 4.0 (Research Genetics, Huntsville, AL) to fit within the marker ladder. Each GROUP was constructed to avoid overlap between markers within SETS but to allow overlap between SETS.

20           The autoradiographic image produced by many markers varied depending on whether the forward or reverse primer was labelled (see Figure 8). Therefore, both primers from each marker were evaluated for image clarity and the ability to distinguish the most intense product(s) or alleles. The appropriate primer was then selected for further use. Optimization of the PCR conditions for each marker was also accomplished using radiolabeling. The strategy of

developing a ladder of markers warranted that the conditions for PCR eliminate nonspecific annealing and background bands. When nonspecific annealing could not be eliminated by raising the annealing temperature, a new marker was chosen for use. Thus uniform PCR conditions as described in Example 1 were used for all the markers chosen except that the annealing temperature was specific to each marker. GROUPS 1 and 2 have 6 and 9 different annealing temperatures, respectively (see Figures 7A and B). An entire microtiter plate containing DNA from a number of different individuals will usually be amplified for a given marker at one temperature at a time, so this should not reduce the overall efficiency of the protocol. For studies with fewer samples a thermocycler block may be used with a lower capacity.

Variability among thermocycler operating temperatures may require adjusting the annealing temperature when switching from one machine to another. Therefore the use of the protocols described for marker GROUPS 1 and 2 should be preceded by a reevaluation of the suggested annealing temperatures for optimal performance. This can generally be carried out once on a few markers and when necessary the annealing temperatures can be adjusted up or down for all the markers for that machine.

The intensity of the products varied considerably from marker to marker. When markers were radiolabelled and a SET was run on the same gel, detecting all of the products on the gel with a single film exposure was often impossible. Attempts to score on a single gel the larger products in each SET using radioactive-based techniques were unsuccessful. Although gradient gels improved the band spacing, a maximum of 4-5 markers could be resolved per gel on autoradiographs. An autoradiograph of GROUP 2 SET B is shown in Figure 9. The range of intensity in the products of this SET is typical of this type of marker and multiple autoradiographs are required for genotyping. These problems are partially overcome by the use

of fluorescent labels (Ziegle *et al.*, 1992). Fluorescent signal detection is linear over a greater range, so that the markers with the weakest product intensity are more readily typed in real-time along with the most intense products from other markers.

Marker GROUPS 1 and 2 are described in Figures 7A and B, respectively. The primers  
5 sequence, chromosomal location, choice of labelled primer, and optimal annealing temperature is listed for each locus. GROUP 1 is composed of a combination of 21 di-, tri-, and tetranucleotide markers from multiple linkage groups. The product sizes range from 66 to 322 nucleotides. Group 2 is composed of 24 dinucleotide markers with products ranging in size from 75 to 349 nucleotides. The mean heterozygosity for both GROUPS is 74%.

10 Scoring of the fluorescent products using the ABI 373 sequencer and GeneScan 672 software was unambiguous in samples that were desalted by ethanol precipitation. Desalting was carried out as follows: 5  $\mu$ l of each PCR product from the same SET (like color) was combined. Then 1.0  $\mu$ l per marker per SET was combined for each of the 3 SETS giving a final volume equal to the total number of markers in the GROUP. Sample handling was otherwise exactly  
15 as described above for the individual fluorescent markers.

A typical set of electrophoretograms of each SET from GROUP 2 for a single individual is illustrated in Figure 5. Each of the alleles can be easily recognized by the unique signature of the stuttering bands for these dinucleotide repeat markers amplified by PCR. Samples that were not desalted were difficult to score because the mobilities of the products and the  
20 ROX-2500 internal lane standards were altered. Salt and primer loads become a problem when combining multiple products for electrophoresis because the necessary volume reduction results in sample concentration. The salt concentration rises with the product concentration and

- 43 -

interferes with the separation of the products and standards. This becomes critical when pooling 21 to 24 markers.

5 It will be understood that while the invention has been described in conjunction with specific embodiments thereof, the foregoing description and examples are intended to illustrate, but not limit the scope of the invention. Other aspects, advantages and modifications will be apparent to those skilled in the art to which the invention pertains, and these aspects and modifications are within the scope of the invention, which is limited only by the appended claims.

**CLAIMS:**

1. A kit for use in automated genotyping within a population comprising at least 4 GROUPS of at least three SETS each comprising labelled pairs of primers for amplification of DNA by polymerase chain reaction (PCR),

5 each primer pair having unique sequence found in the flanking sequences of a microsatellite sequence comprising a nucleotide repeat sequence flanked by unique sequences, such that a polymerase chain reaction (PCR) primed with the primer pair amplifies the nucleotide repeat sequence and at least some immediately adjacent unique sequences of the microsatellite sequence to produce a PCR product identified with the primer pair, wherein the  
10 microsatellite sequences are nucleotide repeat sequences that are polymorphic within the population,

each SET consisting of at least 6 primer pairs, each primer having the sequence of unique sequences respectively flanking at least 6 microsatellite sequences in the genome, such that the length of the segment amplified by a particular primer pair differs from the length of  
15 all other segments in the SET by at least 5 nucleotides, and at least one primer of each primer pair is labelled with a fluorescent label that is the same fluorescent label for all primer pairs in the SET,

each GROUP consisting of at least three SETS of primer pairs labelled with fluorescent labels, wherein the wavelength at which the respective fluorescent labels fluoresce  
20 is substantially different for the labelled primers in each of the respective SETS,

wherein the distance in the genome between one microsatellite sequence amplified by a primer pair of the kit and the nearest other microsatellite sequence amplified by another primer pair of the kit is at least 2 centimorgans (cM) and no more than 50 cM.

2. The kit of claim 1, wherein the PCR products identified with any primer pair amplifying microsatellite sequences containing dinucleotide repeats differ in length from PCR products identified with all other primer pairs of the same SET by at least 9 nucleotides.

3. The kit of claim 1, wherein one of said GROUPS consists of the three SETs of Figure 7A.

4. The kit of claim 1, wherein one of said GROUPS consists of the three SETs of Figure 7B.

5. The kit of claim 1, containing the 6 SETs shown in Figures 7A and 7B.

6. A method of analyzing genomic DNA for the presence of polymorphisms comprising

a) extracting DNA from a human sample;

b) combining, in a polymerase chain reaction (PCR) vessel, an aliquot of said DNA from a human sample, at least one primer pair selected from a GROUP in the kit of claim 1, and PCR amplification enzymes;

c) cycling the temperature of each PCR vessel so that PCR products identified with said at least one primer pair are produced by PCR amplification of segments from said DNA from a human sample, each vessel being cycled at an annealing temperature wherein non-specific annealing of the primers to said DNA from a human sample is minimized;

d) then combining all PCR products from all PCR vessels containing primer pairs from one GROUP into a mixture, and subsequently separating the mixture of PCR products electrophoretically by size;

e) detecting separated PCR products by fluorescence detection at wavelengths corresponding to the fluorescent wavelength for each of the fluorescent labels in the kit.

- 46 -

7. The method of claim 6, wherein the step of combining amplified DNA further comprises:

i) contacting each vessel with a plurality of paramagnetic beads carrying on the surface a protein which specifically binds biotin, further wherein one primer of each primer pair is labelled with a fluorescent label and the other with biotin, for a period sufficient for said protein to bind biotin;

ii) separating the magnetic beads from the PCR reaction medium;

iii) separating the two strands of the amplified DNA segments and combining the strands labelled with a fluorescent label for all primer pairs from one GROUP into a mixture.

8. The method of claim 6, wherein the step of combining amplified DNA from the PCR vessels further comprises:

i) contacting each vessel with a plurality of magnetic beads carrying DNA complementary to the sequence of one primer of the primer pair in the vessel for a period sufficient to allow annealing between the primer and the DNA on the magnetic beads;

ii) separating the magnetic beads from the PCR reaction medium; and

iii) eluting the PCR product from the magnetic beads.

9. The method of claim 6, wherein each primer pair of said kit is added to a different PCR vessel in step (b), such that the annealing temperature for temperature cycling in step (c) is the temperature wherein non-specific annealing of the unique primer pair is minimized and PCR product from all PCR vessels containing at least one primer pair from the same GROUP are combined in a single mixture before electrophoretic separation.



10. A method for selecting a SET of PCR primers for use in automated genotyping comprising

selecting at least 6 microsatellite sequences in the human genome, wherein the microsatellite sequences are selected from dinucleotide, trinucleotide and tetranucleotide repeat sequences that are flanked by unique sequences, said microsatellite sequences being separated  
5 from each other by at least 2 centimorgans in the genome and being polymorphic within the population;

constructing primer pairs for each microsatellite sequence, said primers having the sequence of the unique sequences flanking the microsatellite sequences, such that the length  
10 of all polymorphs of the DNA segment amplified by a particular primer pair is detectably different from the length of all polymorphs of other segments amplified by primers in the SET.

11. A kit for use in automated genotyping comprising at least 4 GROUPS of at least 3 SETS of PCR primers obtained by the method of claim 10.

12. The kit of claim 11, wherein at least one primer of each primer pair in the SET  
15 is labelled with a fluorescent label that is the same fluorescent label for all primer pairs in the SET.

13. The kit of claim 11, wherein the length of all polymorphs of the DNA segment amplified by any primer pair amplifying microsatellite sequences containing dinucleotide repeats differs in length from the DNA segment amplified by all other primer pairs of the same SET by  
20 at least 9 nucleotides.

14. A method of analyzing genomic DNA for the presence of polymorphisms comprising

a) extracting DNA from a human sample;

b) combining, in a polymerase chain reaction (PCR) vessel, an aliquot of said DNA from a human sample, at least one primer pair selected from a GROUP in the kit of claim 11, and PCR amplification enzymes;

5 c) cycling the temperature of each PCR vessel so that PCR products consisting essentially of amplified DNA segments labelled with detectable labels are produced by PCR amplification and the PCR products for all primer pairs in the SET are detectably labelled with the same label, each vessel being cycled at an annealing temperature wherein non-specific annealing is minimized;

10 d) separating electrophoretically by size a mixture containing all PCR products amplified from said DNA from a human sample by any primer pair of said SET;

e) detecting separated detectably labelled PCR products and characterizing them by length.

15 15. The method of claim 14, wherein the mixture in step (d) containing all PCR products amplified from said DNA from a human sample by any primer pair of said SET is obtained by:

i) contacting each vessel with a plurality of paramagnetic beads carrying on the surface a protein which specifically binds biotin, further wherein one primer of each primer pair is labelled with a fluorescent label and the other with biotin, for a period sufficient for said protein to bind biotin;

20 ii) separating the magnetic beads from the PCR reaction medium;

iii) separating the two strands of the amplified DNA segments and combining the strands labelled with a fluorescent label for all primer pairs from one GROUP into a mixture.

16. The method of claim 14, wherein the mixture in step (d) containing all PCR products amplified from said DNA from a human sample by any primer pair of said SET is obtained by:

- i) contacting each vessel with a plurality of magnetic beads carrying DNA complementary to the sequence of one primer of the primer pair in the vessel for a period sufficient to allow annealing between the primer and the DNA on the magnetic beads;
- ii) separating the magnetic beads from the PCR reaction medium; and
- iii) eluting the PCR product from the magnetic beads.

17. A kit for analysis by polymerase chain reaction (PCR) of a genomic region containing at least 6 known loci at which genetic rearrangement is diagnostic for a disease, comprising at least one SET containing at least 6 PCR primer pairs,

each primer pair having the sequence of unique sequences flanking one of said at least 6 loci of genomic rearrangement, such that a polymerase chain reaction (PCR) primed with the primer pair amplifies the DNA segment surrounding the locus of rearrangement to produce a PCR product of characteristic length, wherein the length of the PCR product is associated with specific diagnostic information, and wherein the length of the PCR product amplified by a particular pair of primers differs from the length of all other PCR products amplified by other primers in the SET and the PCR products for all primer pairs in the SET are detectably labelled with the same label.

18. A diagnostic method for detection by polymerase chain reaction (PCR) of genomic rearrangement in a genomic region containing at least 6 known loci at which genetic rearrangement is diagnostic for a disease, comprising

- (a) extracting DNA from a human sample;

- 50 -

(b) combining, in a polymerase chain reaction (PCR) vessel, an aliquot of said DNA from a human sample, at least one pair of amplification primers selected from a SET of at least 6 primer pairs, and PCR amplification enzymes, each primer pair of said SET having the sequence of unique sequences flanking one of said at least 6 loci of genomic rearrangement, such that a polymerase chain reaction (PCR) primed with the primer pair amplifies the DNA segment surrounding the locus of rearrangement to produce a PCR product of characteristic length, wherein change in the length of the PCR product is associated with rearrangement at the locus of rearrangement, and wherein the length of PCR products amplified by a particular pair of primers differs from the length of all other PCR products amplified by other primers in the SET;

c) cycling the temperature of each PCR vessel so that PCR products consisting essentially of amplified DNA segments labelled with detectable labels are produced by PCR amplification and the PCR products for all primer pairs in the SET are detectably labelled with the same label, each vessel being cycled at an annealing temperature wherein non-specific annealing is minimized;

d) separating electrophoretically by size a mixture containing all PCR products amplified from said DNA from a human sample by any primer pair of said SET;

e) detecting separated detectably labelled PCR products and characterizing them by length.

19. The method of claim 14, wherein each primer pair of said SET is added to a different PCR vessel in step (b), such that the annealing temperature for temperature cycling in step (c) is the temperature wherein non-specific annealing of the unique primer pair is minimized

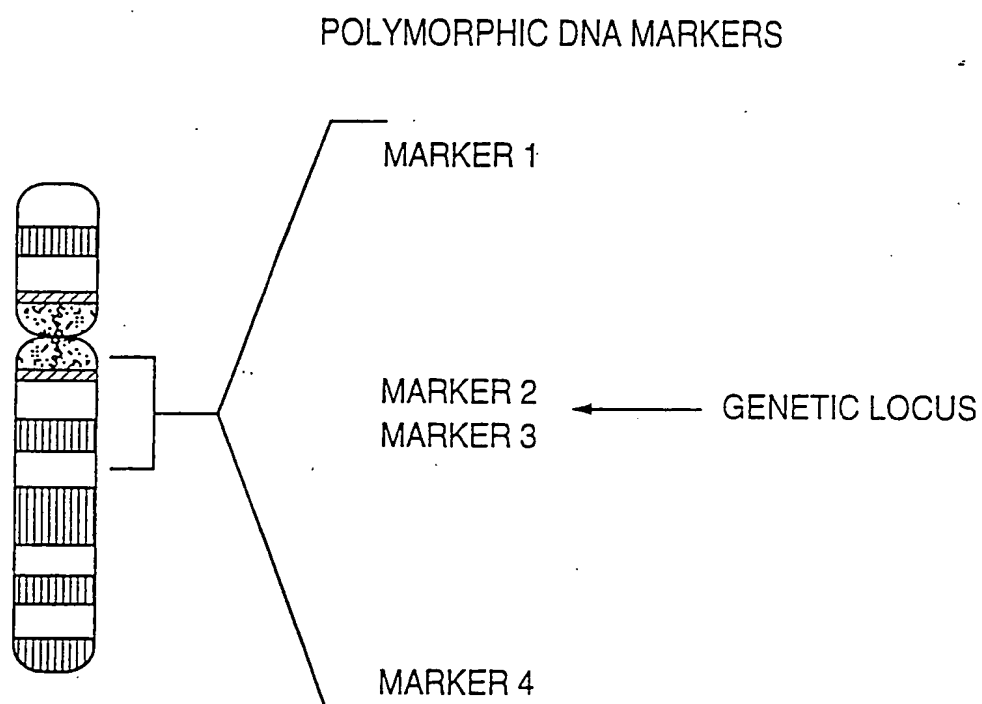
- 51 -

and PCR product from all PCR vessels containing at least one primer pair from said SET are combined in a single mixture before electrophoretic separation.

SUBSTITUTE SHEET (RULE 26)

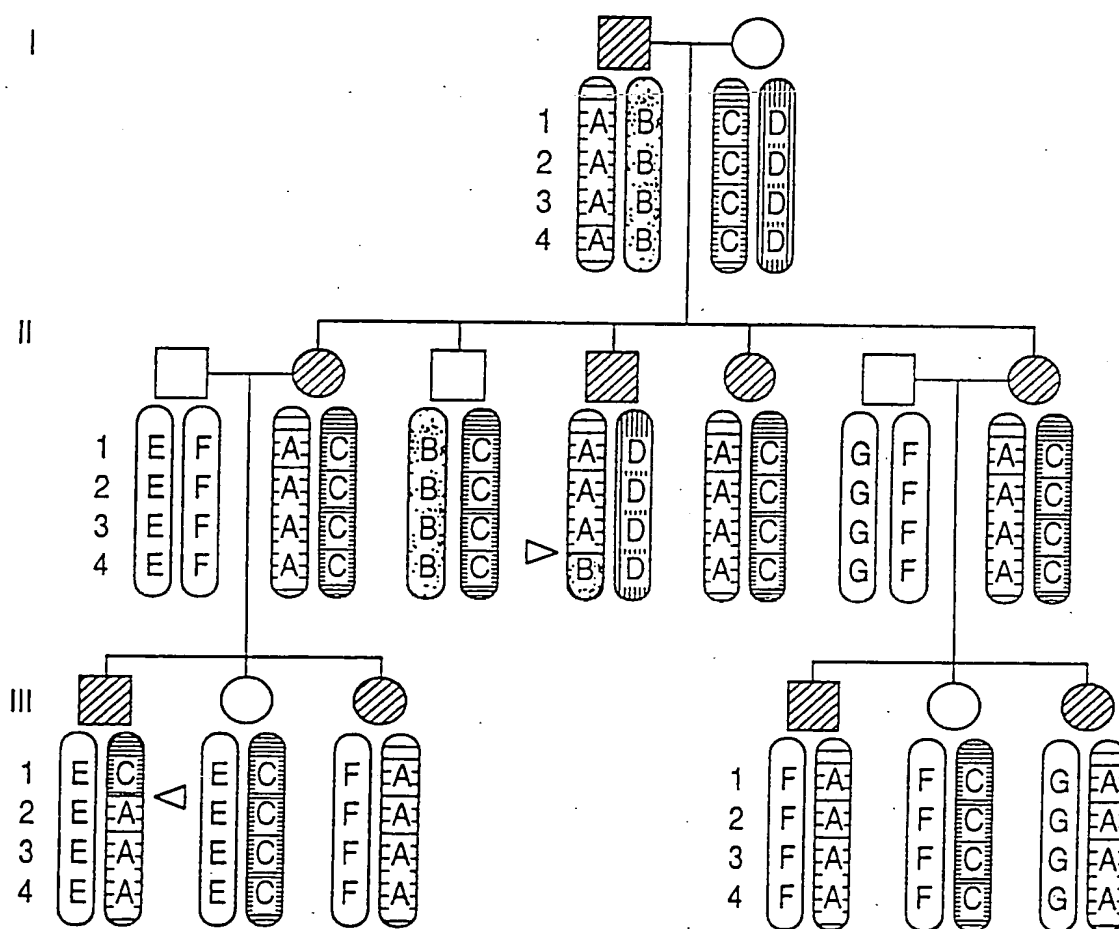
1/50

**FIG. 1**



2/50

FIG. 2



3/50

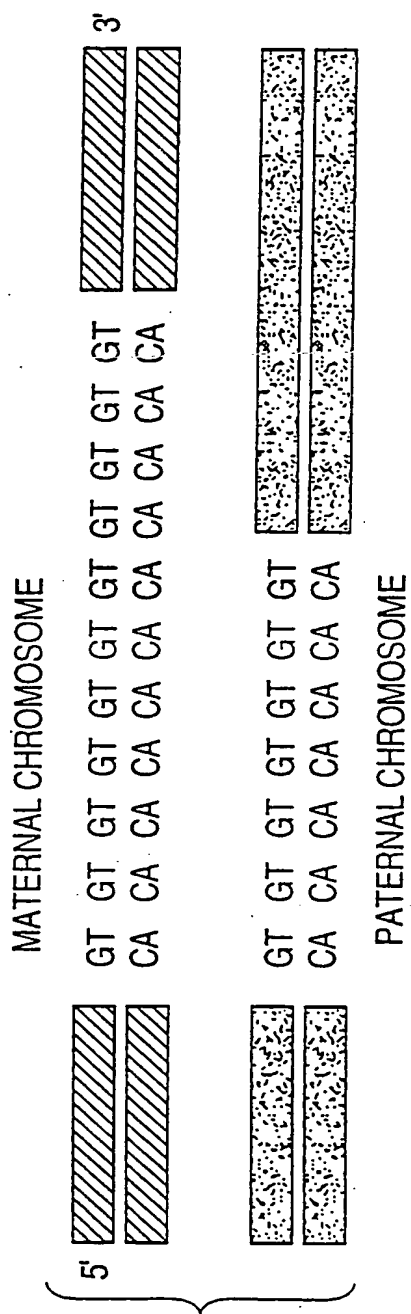
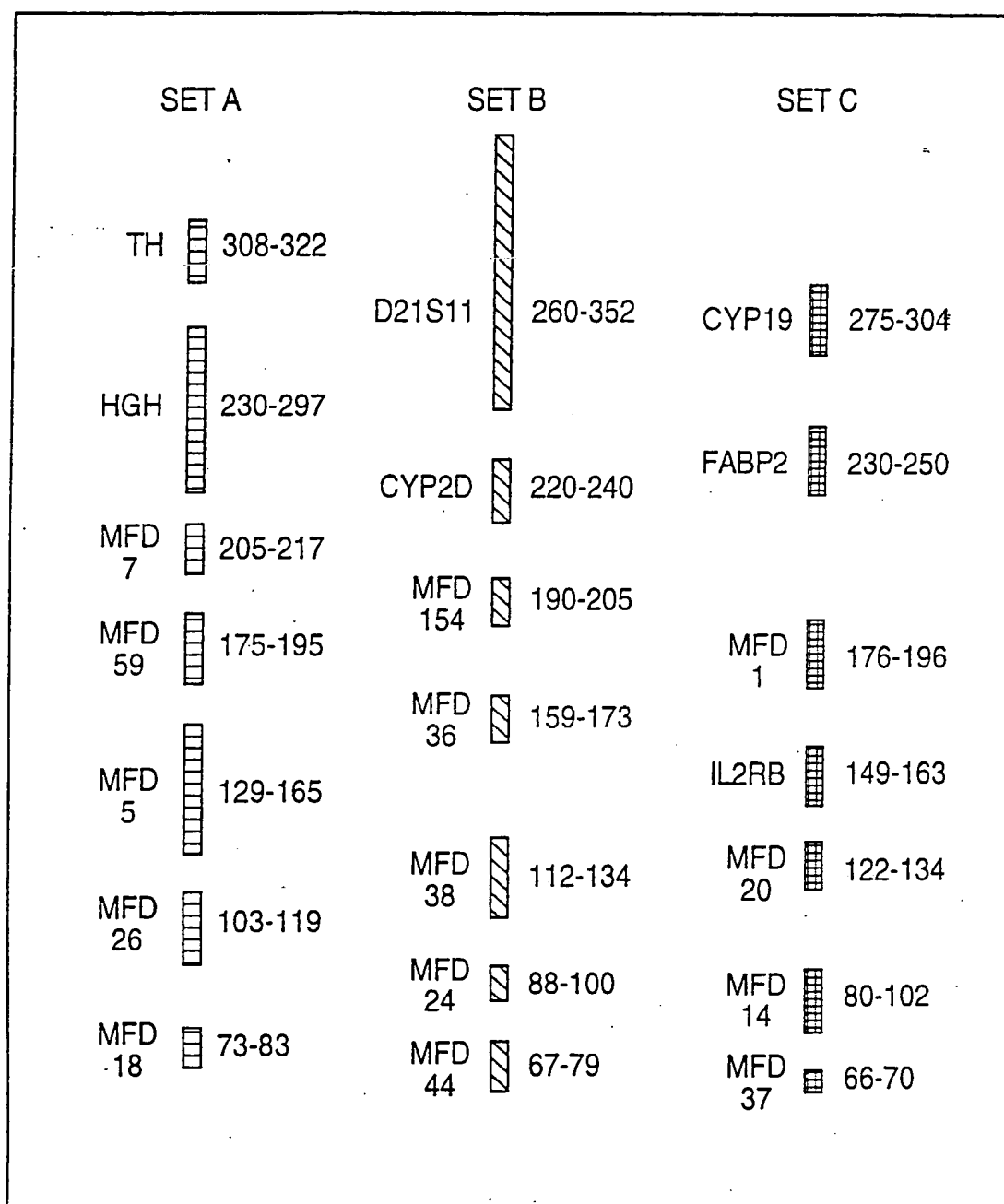
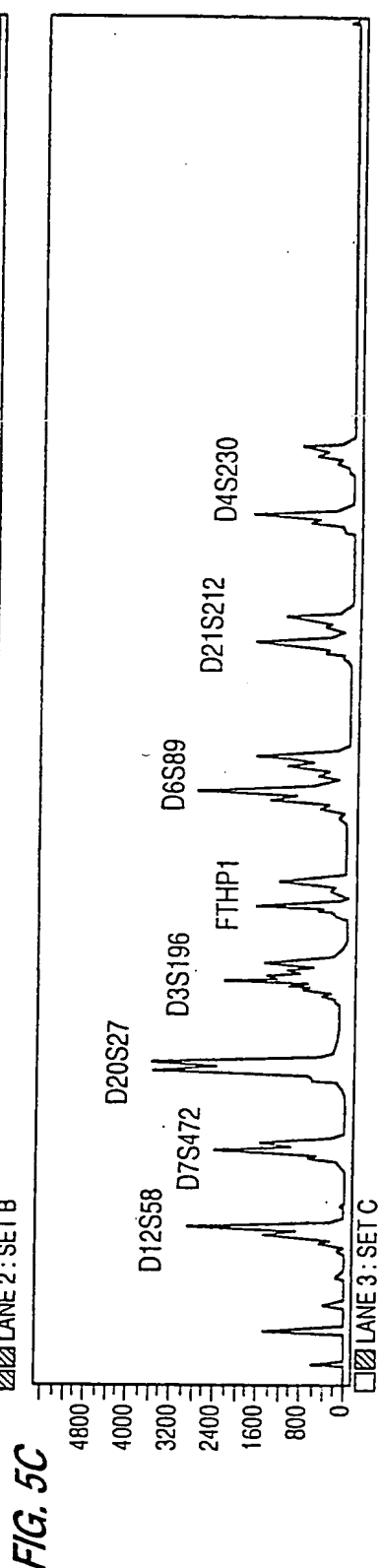
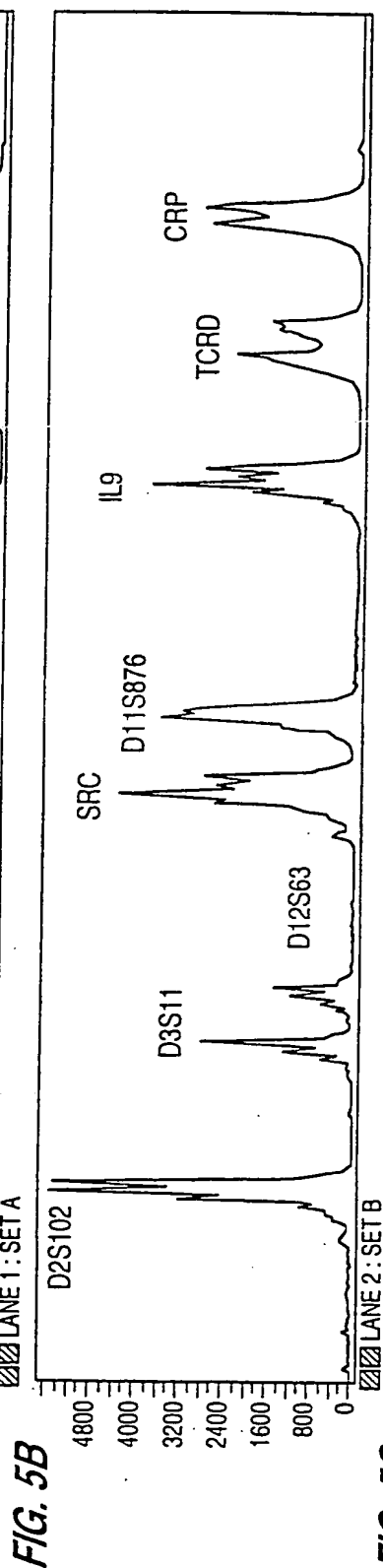
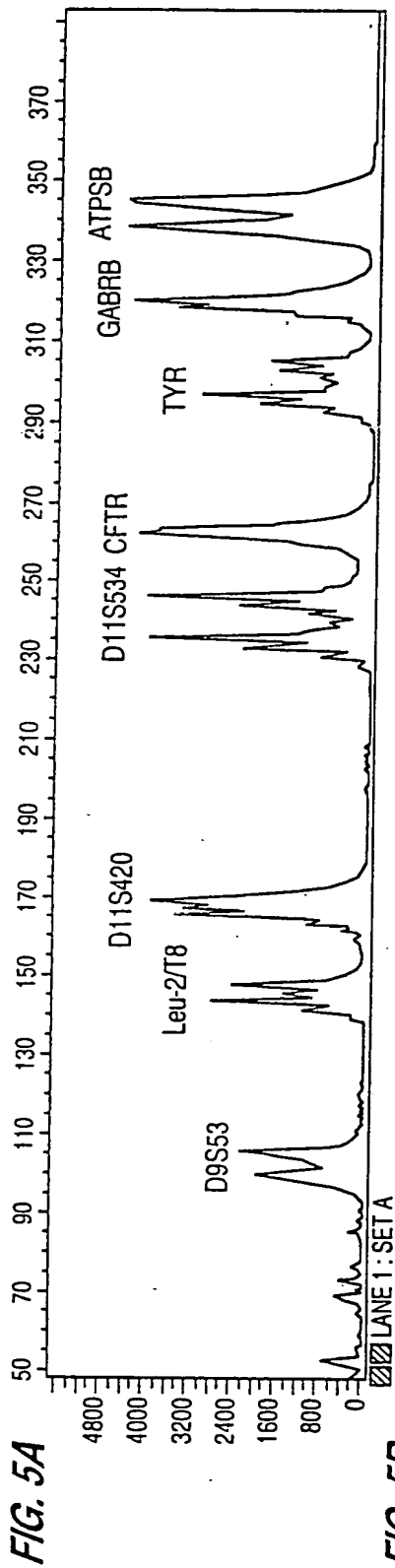


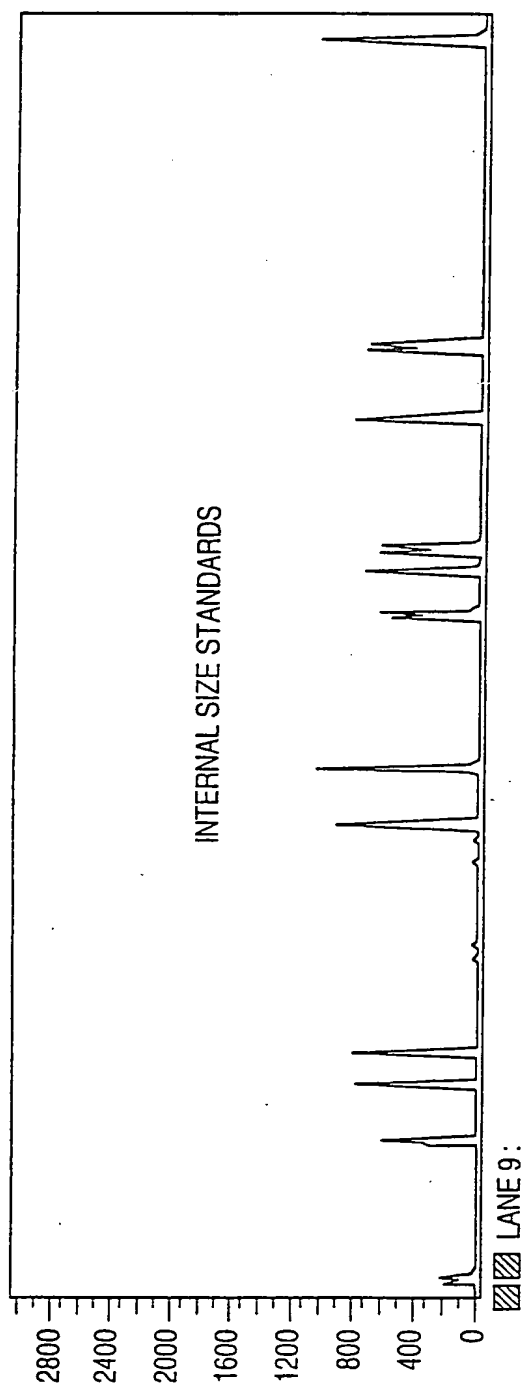
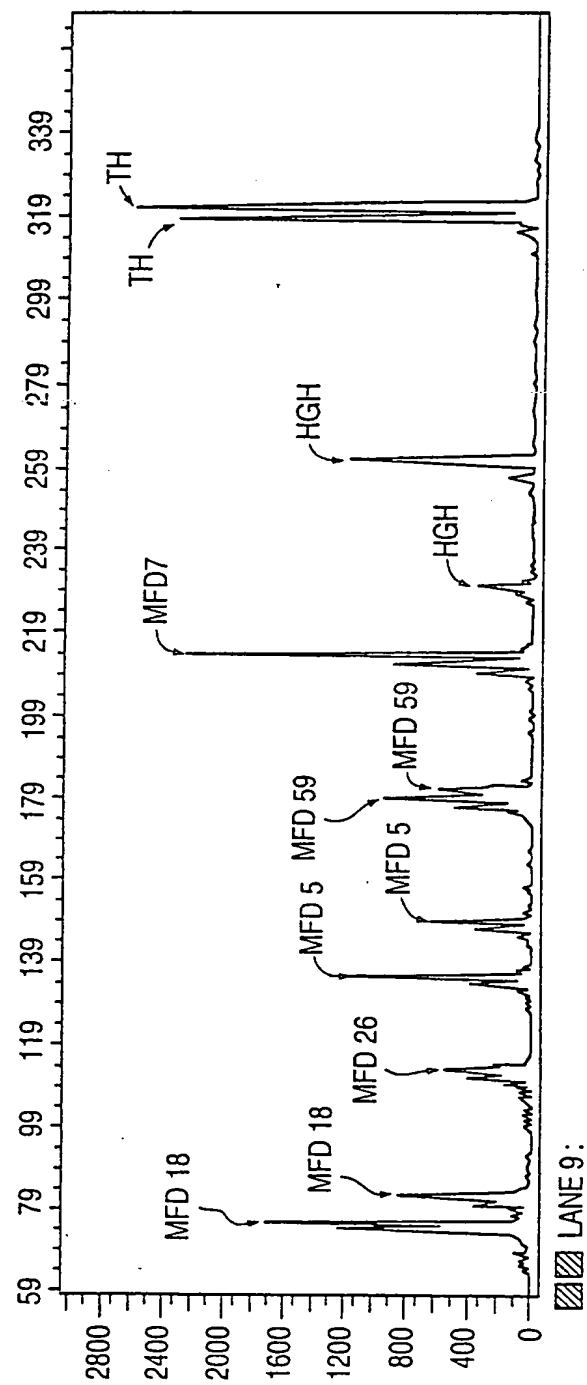
FIG. 3



4/50

**FIG. 4**





7/50

**FIG. 7A-1**

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
TH	(308-322)	75%	11
HGH	(229-297)	83%	17
D10S88	(205-217)	54%	10
D4S174	(175-195)	98%	4
APOC2	(129-165)	80%	19
D18S34	(103-119)	79%	18
D8S85	(74-84)	79%	8
<b>SET B</b>			
D21S11	(260-352)	82%	21
CYP2D	(220-240)	70%	22
D5S211	(186-204)	72%	5
D2S72	(159-173)	71%	2
D4S175	(112-134)	82%	4
D16S261	(88-100)	67%	16
D13S71	(67-79)	76%	13
<b>SET C</b>			
CYP19	(275-304)	91%	15
FABP2	(230-250)	64%	4
IGF1	(176-196)	54%	12
IL2RB	(149-163)	91%	22
D7S435	(122-134)	59%	7
D9S43	(80-102)	83%	9
D19S76	(66-70)	52%	19

8/50  
**FIG. 7A-2**  
**GROUP 1**

A Primer	B Primer
5'-GTC AGC ACC CCA ACC AGC CT-3'	5'-ACC GAA GAC CCC TCC TGT GG-3'
5'-TCC AGC CTC GGA GAC AGA AT-3'	5'-AGT CCT TTC TCC AGA GCA GGT-3'
5'-GTT AGC ATA ATG CCC TCA AG-3'	5'-CGA TGG AGT TTA TGT TGA GA-3'
5'-AAG AAC CAT GCG ATA CGA CT-3'	5'-CAT TCC TAG ATG GGT AAA GC-3'
5'-CAT AGC GAG ACT CCA TCT CC-3'	5'-GGG AGA GGG CAA AGA TCT AT-3'
5'-CAG AAA ATT CTC TCT GGC TA-3'	5'-CTC ATG TTC CTG GCA AGA AT-3'
5'-AGC TAT CAT CAC CCT ATA AAA T-3'	5'-AGT TTA ACC ATG TCT CTC CCG-3'
5'-CTG TTA TGG GAC TTT TCT CA-3'	5'-AAT GTA TGA AGT GGT ATG AT-3'
5'-ATG ACT TCC CCA CTT TTT AC-3'	5'-GCT GAG ATG GGA GGA TTG CT-3'
5'-ACT TTG AAA ACC ACT GGC CT-3'	5'-ATG TAT CTA GCC ATG GTA GC-3'
5'-AGC TAT AAT TGC ATC ATT GCA-3'	5'-TGG TCT ATA ACT GGT CTA TG-3'
5'-ATC TCT GTT CCC TCC CTG TT-3'	5'-CTT ATT GGC CTT GAA GGT AG-3'
5'-AAG CTT GTA TCT TTC TCA GG-3'	5'-ATC TAC CTT GGC TGT CAT TG-3'
5'-GTA TTT TTG GTA TGC TTG TGC-3'	5'-CTA TTT TGG AAT ATA TGT GCC T-3'
5'-AAT CTT CTT TTT TGT CTA TGA-3'	5'-CGT TTG ACT CCG TGT GTT TGA-3'
5'-GTG CCA TTT TAC AGT CTC CT-3'	5'-TTT CCA TTG TCT GTC CGT TT-3'
5'-GCT AGC CAG CTG GTG TTA TT-3'	5'-ACC ACT CTG GGA GAA GGG TA-3'
5'-GAG AGG GAG GGC CTG CGT TC-3'	5'-CAC CCA GGG CCA GAT AAA GA-3'
5'-TTA AAA TGT TGA AGG CAT CTT C-3'	5'-TTT GAG TAG GTG GCA TCT CA-3'
5'-TTC TGA TAT CAA AAC CTG GC-3'	5'-AAG GAT ATT GTC CTG AGG A-3'
5'-AAA AGT GTG TTA CTT TCA GAA C-3'	5'-ACA AGG TGA CAA GGT GCC TA-3'

9/50

**FIG. 7A-3**

Annealing Temperature	Labeled Primer
62°C	B
62°C	A
54°C	B
58°C	A
66°C	A
58°C	A
54°C	A
54°C	B
56°C	B
58°C	B
58°C	B
58°C	B
62°C	B
58°C	A
58°C	B
60°C	A
62°C	B
66°C	A
58°C	B
58°C	A
58°C	B

10/50

**FIG. 7B-1**

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
ATPSB	(337-343)	60%	12
GABRB	(310-318)	72%	4
TYR	(286-298)	58%	11
CFTR	(258-276)	82%	7
D11S534	(228-244)	74%	11
D11S420	(188-208)	66%	11
Leu-2/T8	(138-170)	71%	2
D9S53	(93-127)	87%	9
<b>SET B</b>			
CRP	(331-349)	56%	1
TCRD	(309-319)	74%	14
IL-9	(271-283)	63%	5
D11S876	(216-242)	89%	11
SRC	(193-207)	71%	20
D12S63	(161-175)	72%	12
D3S11	(135-147)	93%	3
D2S102	(102-126)	86%	2
<b>SET C</b>			
D4S230	(276-302)	83%	4
D21S212	(240-260)	86%	21
D6S89	(199-227)	88%	6
FTHP 1	(171-181)	91%	6
D3S196	(149-161)	68%	3
D20S27	(128-138)	65%	20
D7S472	(104-116)	70%	7
D12S58	(75-91)	61%	12

11/50  
**FIG. 7B-2**  
**GROUP 2**

**A Primer****B Primer**

5'-AAA CCC AAA CCC AGA GGA TT-3'  
 5'-GGC ATG TCA TTT TCG TAA GC-3'  
 5'-AAT ATG GCT ACA GCA TTG GA-3'  
 5'-GAG CGA CAG CAA AAT CAG CC-3'  
 5'-ATA TGG AAA CTC TCC GTA CT-3'  
 5'-AGT TAC ACC GGT TCT GCA GA-3'  
 5'-ACT GCC TCA TCC AGT TTC AG-3'  
 5'-TCC TGG CTT TAA ACT TCA CAC AC-3'

5'-GAA CAG AAC AGT GGA GCA TC-3'  
 5'-TAG GAG GCA GAG GAT GGT TC-3'  
 5'-CCC CAC TCT TAG CCA TTG TA-3'  
 5'-TGG AGA TGT GCC ATA GAG GT-3'  
 5'-TTC AAG TGG TTG CCT CTG GC-3'  
 5'-ATG CTT TAT CCA GAG AAA AG-3'  
 5'-CAA ACT TTC CAC AGT ATC GTT C-3'  
 5'-CCA AAT GCT GGA GAC AGA GAG AA-3'

5'-TTC TCA CAA AGT CAC CAC AT-3'  
 5'-GGC CTC CTG GAA TAA TTC TC-3'  
 5'-CTT GTT CAT CTG CCT TGT GC-3'  
 5'-ATC AAT GGA AAA ATG GGT AA-3'  
 5'-ACT GGG GAA CAT GGT GGG GT-3'  
 5'-TTT ATG CGA GCG TAT GGA TA-3'  
 5'-TCC TCA AAA TGA AGA ACA CA-3'  
 5'-CCT GGA AAA ATG GCT CAC C-3'

5'-AGG TGG GTG GAT AAC TTG AG-3'  
 5'-GTG GGC CAC ATT AGG AAC AG-3'  
 5'-TGG GCG ATT TGT TCA TTG TG-3'  
 5'-TGG AAG GAC GGG AAA TAA TA-3'  
 5'-GCA ACC ATG GAG AGT CTG GA-3'  
 5'-GAT TAA TGA TAG TGC TAT CC-3'  
 5'-GAG CAG GCA CTT GTT AGA TG-3'  
 5'-GGA ATA TGT TTT TAT TAG CTT GT-3'

5'-GGC ATA CGA GAA AAT ACT GT-3'  
 5'-CAC CAG CCC CAT TCC TTA GC-3'  
 5'-GAG ACA CAG AGC AAA TAG GT-3'  
 5'-TCA GGA AAA CTG CCT GAG G-3'  
 5'-AGC AAC TTG CCC AGG CTA TGA-3'  
 5'-CAT CAT TAA TTG GAT TGT GG-3'  
 5'-GTT TCC TTG AGA AGA ATG GAG C-3'  
 5'-ACC CCT CCC TCC CTC CAT CAC AC-3'

5'-TAG GGA AAA TGA CAG GAA AA-3'  
 5'-CAT TTT AAT GAA CAC CGC TC-3'  
 5'-ACC TAA GCG ACT GCC TAA AC-3'  
 5'-TAT CTT TCT CTG TCT GCC TT-3'  
 5'-ATG ATG ATT GCC AAA GGG AA-3'  
 5'-CAC CAC CAT TGA TCT GGA AG-3'  
 5'-AAA AGT CTA GTG TTG AGT GT-3'  
 5'-GGA AAA TCA GTC TCT AGT TG-3'



12/50

**FIG. 7B-3**

Annealing Temperature	Labeled Primer
65°C	A
68°C	A
63°C	B
58°C	B
63°C	B
60°C	B
65°C	A
64°C	A
62°C	A
60°C	B
64°C	B
62°C	B
66°C	A
62°C	A
66°C	A
68°C	A
64°C	A
58°C	B
60°C	A
54°C	A
65°C	A
58°C	A
62°C	A
60°C	B

13/50

**FIG. 7C-1**

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D1S198	(308-322)	81%	1
D1S244	(285-296)	82%	1
D1S223	(252-264)	77%	1
D1S234	(226-238)	82%	1
D1S201	(186-204)	73%	1
D1S243	(142-170)	87%	1
D1S197	(115-129)	80%	1
D1S226	(90-106)	82%	1
<b>SET B</b>			
D7S527	(273-297)	75%	7
D7S531	(241-255)	77%	7
D7S529	(218-226)	68%	7
D1S215	(189-207)	73%	1
D1S305	(156-176)	83%	1
D1S250	(133-147)	77%	1
D2S157	(104-122)	75%	2
D1S255	(74-88)	76%	1
<b>SET C</b>			
D1S190	(293-331)	95%	1
D1S196	(267-279)	75%	1
D1S220	(231-251)	83%	1
D1S192	(203-211)	67%	1
D4S424	(178-192)	84%	4
D1S229	(150-166)	77%	1
D1S225	(111-113)	79%	1
D1S202	(77-91)	76%	1

14/50

# FIG. 7C-2

## GROUP 3

## A Primer

## B Primer

5'-GAC TTC ACC ATC AAC GCC TG-3'

5'-GAG CAG CAC CGT ACA AAT-3'

5'-TAA CAT GAG CGA ATG GAC AA-3'

5'-GCC CAG GAG GTT GAG G-3'

5'-GGT ATG GAA GTC ACC CAA CA-3'

5'-CAC ACA GGC TCA CAT GCC-3'

5'-TCA TGT CCC TCC TCC CAA AG-3'

5'-GCT AGT CAG GCA TGA GCG-3'

5'-CAG GAA AGT GGA TGT GAC GA-3'

5'-AGC TCC GCT CCC TGT AAT-3'

5'-CAA GGT TTC ACC ACA GTT CT-3'

5'-AAG GCA GGC TTG AAT TAC AG-3'

5'-CTC AAA ATG ACT GAT GGG GT-3'

5'-GCT CCA GCG TCA TGG ACT-3'

5'-GAG CAA GCA TCC AAA AAC GA-3'

5'-GGT CAC TTG ACA TTC GTG G-3'

5'-CAT TGC AAA CTC AGG AGA TA-3'

5'-AAA CTG TGG TCC TGG CTG-3'

5'-AAA TTC TAG ACA TCG CCT GTA A-3'

5'-GAC ACA GGT AGG TTA GAA GGA TG-3'

5'-CCA GNC TCG GTA TGT TTT TAC TA-3'

5'-AAA AAC GTA CTG CCA CAT TC-3'

5'-AGC CAG CAT TAC CTC TGN TAC C-3'

5'-TTA GCA AAT CCC AAG CAA TA-3'

5'-TAA CAG AGG CAT GAA AAC CA-3'

5'-AAA CTA GAG TCC TGG CCT GA-3'

5'-GGT ACC ATC ACC ACA ATC AA-3'

5'-TGT CTT GGT GAA TTG ACC CT-3'

5'-CTG AAA CCT CTG TCC AAG CC-3'

5'-ACT TGT AGG CCT GTT CTG AG-3'

5'-GAT CAC AGA TAT TGG CCC ATA G-3'

5'-GTG ATG GTG GTA AAG GCA GA-3'

5'-GGT GCC AGA CTA TGC AGA CC-3'

5'-GGC TGT GGG TGT TTC TCC TA-3'

5'-GAT CGC CTA TGA CCT CCT TG-3'

5'-TTA ATA AAA ATA CCC CCA CC-3'

5'-GCG CTC TTG GTA TAT GGT ACA G-3'

5'-GAA TGT GAA AGG CTG TGC-3'

5'-TGG CCT GAA TAG ACC ATA AAA A-3'

5'-CAA CAC CCA AAC AGA TGA CC-3'

5'-TAT GCT GAT TTA GGG AGC CC-3'

5'-AGC TCT CAT GNC TTT ACA TTC T-3'

5'-GCT GTC TGT GAG AGT TCG CA-3'

5'-GGA AAT AGG TGT GAA CAA AA-3'

5'-TGT GGG CAA CGT CAC TC-3'

5'-AAA ATT ACA AAG AAG ACC-3'

5'-GCC TGG GTG ACA AAG CA-3'

5'-AGT CTT TCA TGG CCA CTG TG-3'

SUBSTITUTE SHEET (RULE 26)

15/50

FIG. 7C-3

<u>Annealing Temp.</u>	<u>Labeled Primer</u>
----------------------------	---------------------------

64°	R
-----	---

58°	R
-----	---

61°	R
-----	---

62°	F
-----	---

62°	R
-----	---

66°	F
-----	---

60°	F
-----	---

68°	R
-----	---

64°	R
-----	---

68°	R
-----	---

64°	F
-----	---

68°	R
-----	---

70°	R
-----	---

62°	F
-----	---

68°	F
-----	---

64°	F
-----	---

68°	R
-----	---

66°	F
-----	---

70°	F
-----	---

58°	F
-----	---

64°	R
-----	---

52°	F
-----	---

64°	F
-----	---

67°	F
-----	---

16/50

*FIG. 7D-2*  
GROUP 4

---

A Primer

---

5'-GAG GCA GGA GAA TCA CTT-3'

5'-AGA TGA GGG GTA ATG TTG GA-3'

5'-TTC GCT CTT TGA TAG GC-3'

5'-CCC CTT GGA AAA TCA CTG-3'

5'-CCT AAG TAG GCA GTT GGT AT-3'

5'-AAC TTA CAC ATT TGG CCC TG-3'

5'-AAC TGC AAC ATT GAA ATG GC-3'

5'-TGG AAA CTA TGT ATC TTG GAG G-3'

5'-CAT ATG CAT ACC ACA CAC-3'

5'-AGC TCA GAG ACA CCT CTC CA-3'

5'-TCA GCC TGA GTT TTC TTT AT-3'

5'-GGT CTG ATG AAA ATG TTC TCA AGC-3'

5'-AAC GTC TGC TCG TCA GAG TC-3'

5'-GCC TTG GGG GTA AAT ACT CT-3'

5'-TTT TCT TTT TTG CAG TTT ATC C-3'

5'-ATC TTC CAA AAA TGT CAT-3'

5'-GGC CAG GCT TTG TTC AGA-3'

5'-TTT AGC CTG AAA ATA CAC GC-3'

5'-TGC ACA TTA AAG GAA CAG GT-3'

5'-GAT CTG ATT AGT ATT GTC TGC TTG A-3'

5'-AAA TGT GAG TAG AAG GGA TAG GTT-3'

5'-GAG TGG CGG TGA GAA GGT AT-3'

5'-TGG AAT TTC TCC ATG TTG AG-3'

5'-GAA AAG AAT GCT GGA TAG-3'

SUBSTITUTE SHEET (RULE 26)

17/50

*FIG. 7D-3*

---

**B Primer**

---

5'-ATG GTT GTA GAT GAG ACT GG-3'

5'-AAG CAT CTT AAT GGA TGG AAA-3'

5'-ATT TCA TTT GTA ATT TAC TAG CAG-3'

5'-CCA TGA ATA AGC CTT GCC-3'

5'-CAC AGC AGG GGT TCA TTT TT-3'

5'-TCA ATC TGT GGA GTC ATT GG-3'

5'-GGG ACC ATA GTT CTT GGT GA-3'

5'-GCN GGC TTT AGG GTG G-3'

5'-AAT CTT ATT GCT GTC TCA-3'

5'-CTG TAT TAG GAT ACT TGG CTA TTG A-3'

5'-CAA GGA GCA GGA AGA ACA GC-3'

5'-TAG ACT GGG TTG TTA GGG ACT CTC-3'

5'-CGA CTA CGT GCT GGC TAC TT-3'

5'-GGA ATT ACA GGC CAC TGC TC-3'

5'-CAC TTC AGT GCC TTC TTG AGA-3'

5'-CAT AAT AGG AGA ATA AGA-3'

5'-CAG GGT CTA TGA TAC GCT TT-3'

5'-GCT TTG CTC CTA GAG TCC AG-3'

5'-CAT AAT TTG CTG CTT TGG AT-3'

5'-GCT TTA TAG GAG GTA TCT TTN TGT G-3'

5'-TAA AAA AGN CCG ACT AGA CC-3'

5'-AGC CAT TGC TAT CTT TGA GG-3'

5'-AAG AGC TAT GAA AAG AGT TAA AGG A-3'

5'-CCA GTT TTT ATG GAC GGG GT-3'

SUBSTITUTE SHEET (RULE 26)

18/50

*FIG. 7D-4*

Annealing Temp.	Labeled Primer
--------------------	-------------------

64°	R
-----	---

59°	F
-----	---

57°	R
-----	---

59°	F
-----	---

64°	R
-----	---

61°	F
-----	---

80°	F
-----	---

66°	F
-----	---

69°	R
-----	---

64°	R
-----	---

68°	R
-----	---

60°	R
-----	---

50°	R
-----	---

66°	R
-----	---

62°	F
-----	---

58°	R
-----	---

62°	F
-----	---

64°	R
-----	---

66°	F
-----	---

62°	F
-----	---

19/50

**FIG. 7E-1**

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D5S436	(334-354)	85%	5
D4S405	(279-299)	87%	3
D4S431	(246-270)	83%	4
D3S1303	(196-220)	78%	3
D3S1296	(176-186)	74%	3
D3S1271	(146-158)	75%	3
D4S407	(111-135)	87%	4
D4S404	(89-101)	79%	4
<b>SET B</b>			
D3S1279	(264-282)	86%	3
D2S176	(240-250)	70%	2
D2S151	(211-229)	83%	2
D3S1305	(189-198)	74%	3
D4S403	(155-169)	76%	4
D4S411	(135-143)	66%	4
D2S173	(117-125)	70%	2
D4S392	(93-107)	84%	4
<b>SET C</b>			
D5S418	(297-313)	80%	5
D6S313	(279-285)	68%	6
D6S314	(243-259)	81%	6
D6S289	(215-227)	80%	6
D7S513	(173-201)	84%	7
D7S492	(145-155)	78%	7
D7S478	(118-130)	70%	7
D6S294	(86-108)	83%	6



20/50

*FIG. 7E-2*  
**GROUP 5**

---

**A Primer**

---

5'-AGG TCA TTG AGG TTT ATA TTC CCA-3'

5'-ATC AGG AGA TGT TGC CTT GC-3'

5'-AGG CAT ACT AGG CCG TAT T-3'

5'-CAG ACA ATG GCT TCC AAA AGT A-3'

5'-CCT GAA GGG TGT AAT TTT CA-3'

5'-TGA TTG GAG GTG GTA GAG GT-3'

5'-ATA ATA TCC TTT GAT CCT TTC GCT A-3'

5'-TTC CTC ATT TAC CTG CAC TAA G-3'

5'-CAC CAT CTG TGT GGT ATT GG-3'

5'-TTG TGC ACT CGT TAT GAG AA-3'

5'-AAC TAA GAC ACA CAA CCC CG-3'

5'-CTG CTG GAA CTT AAA AGT GC-3'

5'-CAA CAG ATC TCC CAA GGT AG-3'

5'-AGG CTG TCT TGG CAG AAA T-3'

5'-GAG GGC TGT TGA CCC AC-3'

5'-TCG GTA AAC ATT CAT CCA GA-3'

5'-AAA CAA AAT AGC CTT CAA AA-3'

5'-TAG GCC CAA GGA ATT NAA AA-3'

5'-AAA ATG ACT TCT TTG GGT GGG C-3'

5'-TTC GCT GAG ATC ATG CCA C-3'

5'-AGT GTT TTG AAG GTT GTA GGT TAA T-3'

5'-ATC TTG GAT TTA GGG TTG GC-3'

5'-TGT GTC ATT ACG CTT TTC ATC-3'

5'-TGC ATT GTT GTC ATG CCT-3'

21/50

**FIG. 7E-3**

---

**B Primer**

---

5'-GAA CCC TAG GAA GTG AAA TAC AAA A-3'

5'-CAG GGC TAT GAT TGG ATG TC-3'

5'-TTC CCA TCA GCG TCT TC-3'

5'-CAA ACT TAG GGT TGT TCC TCA C-3'

5'-TGA GAA GGT GTG TTA GGG TG-3'

5'-AGC TAT CAT GTA GAA AAG CAG CA-3'

5'-AAA TTT GGT TAT TTT TAA GCA AAC T-3'

5'-TTG CTA AAC CTT GGG TGT GT-3'

5'-GAC CTA TTT TGG TTA ACA ATT TAG A-3'

5'-CTG ATG GAG GTT AAG GCA AG-3'

5'-CCA ATT CAG TGG CAT CTA TG-3'

5'-AGA AAT GAG ATA TTG TTT TCG C-3'

5'-CTC ATA ACT CAA AAC CTC TG-3'

5'-GAT GTA ATC CTG TGC TAT GGC-3'

5'-TTG CCT GGA AAC CTG GTA-3'

5'-TGT CAA AAT GGA CCA ATC AG-3'

5'-GCC TGG TAA GTT GAT AGT GT-3'

5'-TCA TCA TCA CCA CAA ATG CT-3'

5'-GTG GGT AGC AAC ACT GTG GC-3'

5'-AGA CCT TTA GGT TGT TCA TGC TG-3'

5'-ATA TCT TTC AGG GGA GCA GG-3'

5'-GGC TCT GCT CCA TCT TCA TA-3'

5'-TCA AAT GGT TCA GGA GAA AGA-3'

5'-TAA AGT CTC CAT CTT CGA TTG T-3'

22/50

FIG. 7E-4

<u>Annealing Temp.</u>	<u>Labeled Primer</u>
----------------------------	---------------------------

68°	F
-----	---

66°	R
-----	---

62°	F
-----	---

68°	F
-----	---

61°	F
-----	---

60°	F
-----	---

64°	F
-----	---

60°	R
-----	---

66°	R
-----	---

66°	F
-----	---

60°	R
-----	---

69°	F
-----	---

66°	R
-----	---

58°	F
-----	---

60°	F
-----	---

69°	F
-----	---

68°	F
-----	---

68°	F
-----	---

62°	F
-----	---

62°	R
-----	---

66°	F
-----	---

SUBSTITUTE SHEET (RULE 26)

23/50

*FIG. 7F-1*

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D6S286	(315-341)	78 %	6
D7S521	(288-306)	71 %	7
D7S505	(262-278)	70 %	7
D6S301	(221-251)	77 %	6
D7S518	(179-201)	88 %	7
D6S292	(141-161)	83 %	6
D6S264	(108-122)	71 %	6
D6S268	(79-93)	75 %	6
<b>SET B</b>			
D5S412	(287-303)	83 %	5
D5S413	(264-276)	70 %	5
D5S428	(241-255)	77 %	5
D5S419	(204-226)	82 %	5
D5S423	(179-191)	77 %	5
D5S421	(152-170)	83 %	5
D6S273	(130-140)	77 %	6
D5S392	(83-117)	92 %	5
<b>SET C</b>			
D7S517	(341-335)	83 %	7
D8S265	(284-307)	75 %	8
D8S282	(260-272)	73 %	8
D8S272	(192-239)	82 %	8
D7S530	(170-182)	78 %	7
D8S275	(139-157)	76 %	8
D8S255	(107-129)	74 %	8
D7S520	(79-97)	70 %	7

SUBSTITUTE SHEET (RULE 26)

24/50

# *FIG. 7F-2*

## GROUP 6

---

### A Primer

---

5'-TCA CCC CTA ATA CCC AAA AC-3'

5'-AGT GGA CAG TTG GTA TCT CA-3'

5'-ACT GGC CTG GCA GAG TCT-3'

5'-CAC AAT CAT ATG TNC CAA TT-3'

5'-CAG TAG GCA GGG GTG G-3'

5'-AAT TCA CAA GAC ACA ATC TCA G-3'

5'-AGC TGA CTT TAT GCT GTT CCT-3'

5'-CAA CAT ACT GCC TCA AAA-3'

5'-TTC GGC CAA AAA CAG AGT CC-3'

5'-AGT CAC CTT CTC TGT CTC CA-3'

5'-AAC ATC TTA GGG CAT CCT G-3'

5'-ATC TTT TAT TGT GGG GTG CT-3'

5'-CTG GGC AAC AAG AGT GAA AT-3'

5'-TGG AAA TAG AAT CCA GGC TT-3'

5'-GCA ACT TTT CTG TCA ATC CA-3'

5'-GCT ATT CCC ACA AAG GCA-3'

5'-ATC ATG GGA AGT GCG TGG-3'

5'-CTT TCC TGC CAA CCT CTT TC-3'

5'-GGG CAC AGG CAT GTG T-3'

5'-GAG AAC TAA TCC CTT CTG GC-3'

5'-TCC CTA CGT TGC ATT TTA-3'

5'-AAA TCG CTA GAA AAT GTC CA-3'

5'-TTT TGG AAT TTC TAG CCT CC-3'

5'-CAA CAG GTC CAG GCT ATG TC-3'

25/50

**FIG. 7F-3**

---

**B Primer**

---

5'-AAT ATG AAG GGA TGT TGA AT-3'

5'-TGT GAT CAG CCC AGG AAG AG-3'

5'-CAG CCA TTC GAG AGG TGT-3'

5'-ATT AAA TGT GCA TAC GCA AA-3'

5'-GGG TGT GTC TGT GTG ACA AC-3'

5'-AGA ACT AAA GTT GCC TGT TCN TGT A-3'

5'-TTT TCC ATG CCC TTC TAT CA-3'

5'-TAC ACA AAA AGG AGG TCA TT-3'

5'-TGA GAA CTT CCA CAT AGC AG-3'

5'-AGG CCT CAT TCA AAA TCT GT-3'

5'-AAT GAT TTA AAA TAG ATT AGG AGC A-3'

5'-TGC CCA GAC TTC TCA CCT-3'

5'-CAA ATT CCA CAA AGC CGT-3'

5'-TCT ATC GTT AAC TTT ATT GAT TCA G-3'

5'-ACC AAA CTT CAA ATT TTC GG-3'

5'-GGC GGA TCA TTG AGT GC-3'

5'-TAA TTA GTT GCT GGT TTG AA-3'

5'-TTG GGT TCA AGC GAT TCT CC-3'

5'-GGC TGC ATT CTG AAA GGT TA-3'

5'-AGC TTC ATA AAG AGT CTG GAA AAT-3'

5'-TAC CCA GCC AAA CTA TTA-3'

5'-TCA CAC CTG GGA ATT AGA AG-3'

5'-TGA AAC CCA CAG ATA TTG GG-3'

5'-TAT CCA TAC ACA CCA TGC CA-3'

**SUBSTITUTE SHEET (RULE 25)**

26/50

**FIG. 7F-4**

Annealing Temp.	Labeled Primer
--------------------	-------------------

60°	R
-----	---

60°	F
-----	---

62°	R
-----	---

62°	R
-----	---

62°	R
-----	---

64°	R
-----	---

60°	R
-----	---

62°	F
-----	---

66°	R
-----	---

60°	R
-----	---

68°	R
-----	---

56°	R
-----	---

	F
--	---

	F
--	---

	F
--	---

56°	F
-----	---

	F
--	---

	F
--	---

27/50

*FIG. 7G-1*

---

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D10S211	(289-305)	83%	10
D9S163	(271-279)	71%	9
D12S102	(241-259)	78%	12
D10S223	(221-231)	67%	10
D6S271	(166-208)	85%	6
D9S153	(143-155)	77%	9
D9S170	(108-126)	75%	9
D14S79	(79-89)	67%	14
<b>SET B</b>			
D4S402	(287-323)	92%	4
D14S72	(257-271)	83%	14
D13S221	(223-243)	83%	13
D15S165	(184-208)	80%	15
D14S68	(148-172)	89%	14
D14S64	(126-136)	77%	14
D13S175	(69-83)	76%	13
D15S132	(74-88)	76%	15
<b>SET C</b>			
D18S64	(301-321)	73%	18
D18S68	(270-290)	80%	18
D18S66	(244-262)	86%	18
D15S118	(218-230)	76%	15
D16S420	(179-201)	82%	16
D18S59	(148-164)	82%	18
D16S423	(121-139)	75%	16
D18S57	(88-110)	88%	18



28/50

## **FIG. 7G-2**

### **GROUP 7**

---

#### **A Primer**

---

5'-GCT AGG ATT ACA GGC ACA T-3'  
5'-TGC TGC ACA TCT TAG GGA GT-3'  
5'-CTT TGC AGA ACC CAT GAT TAT GA-3'  
5'-AAT TCT GAA GAG GCA AAT CTA A-3'  
5'-AAC AAT TGG GAA ATG GCT TA-3'  
5'-TTA TGG CAG CCC AAA TGG ACT A-3'  
5'-CAG GCA CAC GCA TAC AC-3'  
5'-AGG TTG ATA GAC CAT GGA GAC A-3'

5'-CTT ACT GTG TTG CCC AAG GT-3'  
5'-TGT AAA GTT TTG TAC ATG GTG TAA T-3'  
5'-TAG CCA TGA TAG GAA ATC AAC C-3'  
5'-GTT TAC GCC TCA TGG ATT TA-3'  
5'-GAG AGG TGG TTT TCA GTG GT-3'  
5'-GGG CAA CAC AGT GAG ACT CT-3'  
5'-TAT TGG ATA CTT GAA TCT GCT G-3'  
5'-CTG ATA ATA AAA CCA GGA AGA CAC-3'

5'-TTC TGG AAA TGG ATA CTG GT-3'  
5'-ATG GGA GAC GTA ATA CAC CC-3'  
5'-AGA GCA AGT CCC TGC C-3'  
5'-TCA AAG ACC CAT ATC AAC CA-3'  
5'-ATT TCC TGA GGT CTA AAG CAC CC-3'  
5'-AGC TTC TAT CCA ACA GGG GC-3'  
5'-AAC AGG CTT GAA AGT CTC TGT C-3'  
5'-TTC AGG GTC TTT TGA AGA GG-3'

SUBSTITUTE SHEET (RULE 26)

29/50

*FIG. 7G-3*

---

B Primer

---

5'-AGG CTC CTA CTA CCG TCA C-3'

5'-ACA GCG CTC AGA AAT CAT ATA A-3'

5'-ATT GCC TTG GAG GGC G-3'

5'-AGG AAA ATA TAC ACA ACC CAA G-3'

5'-TAG GTT GTG GTG GGT GTT AC-3'

5'-GCA GAA TGT TGC CCA AAA CTC A-3'

5'-ACT TCA GGA ATA GCC TTT ACC-3'

5'-TTT TAT TGT TAT GTG GCT TTC A-3'

5'-AGC TCT ATG ATT CAT TTC AAG TTT G-3'

5'-TCC TAA CAT TCT GCT ACC CA-3'

5'-GAG ATC GTG CAG CAC TTG T-3'

5'-GGG CAC ACA GTC CCA A-3'

5'-TCA GGG ATA GTT GGT GGG TA-3'

5'-TGG GAT AGA AGC AAC ACA GA-3'

5'-TGC ATC ACC TCA CAT AGG TTA-3'

5'-TAT TGG CCT GAA GTG GTG-3'

5'-TTT GGA TGC ACA GGA AGT TG-3'

5'-ATG CTG CTG GTC TGA GG-3'

5'-CAG CCT CGG AGA AAC G-3'

5'-GTG CTG AAA AGC GAC ACT TA-3'

5'-TTA GGC CCA GTC CAC ACT CAA G-3'

5'-ACC AGA ATG TGA ACG ACC CT-3'

5'-GCC TAT TTG ATA ATG CTG TAC G-3'

5'-AGA AGG CAT TAA ATT TTG CA-3'

30/50

***FIG. 7G-4***

Annealing Temp.	Labeled Primer
--------------------	-------------------

	R
66°	F

62°	F
	R
	F
	R

64°	F
64°	R
64°	F
56°	F
	R
62°	R

	R
--	---

64°	R
	R
	F

	R
--	---

31/50

*FIG. 7H-1*

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D11S914	(275-285)	71%	11
D11S910	(249-261)	71%	11
D17S784	(226-238)	78%	17
D22S274	(202-214)	78%	22
D19S216	(179-191)	76%	19
D21S259	(117-131)	80%	21
D20S103	(92-106)	71%	20
<b>SET B</b>			
D12S89	(254-288)	79%	12
D10S205	(224-244)	90%	10
D12S101	(195-213)	82%	12
D12S91	(176-181)	70%	12
D11S902	(145-163)	81%	11
D10S249	(118-134)	75%	10
D11S903	(99-109)	75%	11
<b>SET C</b>			
D17S801	(258-336)	86%	17
D17S809	(229-247)	72%	17
D20S100	(194-218)	77%	20
D19S213	(174-184)	69%	19
D18S58	(144-160)	74%	18
D18S52	(116-130)	77%	18
D17S793	(95-109)	70%	17

SUBSTITUTE SHEET (RULE 26)

32/50

*FIG. 7H-2*  
GROUP 8

---

A Primer

---

5'-ATC TCA TGG GAG TAC CGT TG-3'  
5'-AGC TTT GCA GAC AAG GCA AG-3'  
5'-GAG TCT CCT AAA TGC TGG GG-3'  
5'-GTC CAG GAG GTT GAT GC-3'  
5'-TCT TGT CAC TCT AAC TCC GC-3'  
5'-AGA ATG TGG TCT CAC AAG CC-3'  
5'-GTT CAT AGA GGG ACA AGA CAC AGT-3'

5'-ATT TGA GAG CAG CGT GTT TT-3'  
5'-GGC ACT TGT AAT CCC CG-3'  
5'-CAA AAA AAT GTT TTA CTA AGC AGG-3'  
5'-TTC ACA ACA GCC AAT GGT AG-3'  
5'-CCC GGC TGT GAA TAT ACT TAA TGC-3'  
5'-AAC TGG TTT TGG TAG TGA GA-3'  
5'-AAC ACT TCG ATG TTC CTT CC-3'

5'-CCT CAA ACC GGA CAA CTA TTT-3'  
5'-CAA AAA GGC AGA ATG CAG TA-3'  
5'-ATT GGG TTT ACT TGT GCC TT-3'  
5'-CCT CCA ATC TGC ACC TGA CT-3'  
5'-GCT CCC GGC TGG TTT T-3'  
5'-TTN CAA CAT AGG TTA TAC GCG-3'  
5'-TGT TGG AGT TAA TGT GCC AT-3'

33/50

*FIG. 7H-3*

B Primer	Annealing Temp.	Labeled Primer
5'-GAC CCA CAT CAC CAT TAC TG-3'		
5'-TCC CTG CTC ATA AGT CAG CC-3'		
5'-AGC TCC TGC ACA GTT CTT AAA TA-3'		
5'-AGT GCC CAT TTC TCA AAA TA-3'		
5'-GGC CCA TGT CTT TTT TAG GT-3'		
5'-AGG GAA TGT CAA TGA AAA CC-3'		
5'-CCA TGA TGT TTG GTT AAT CAC A-3'		
5'-CCA TTA TGG GGA GTA GGG GT-3'		
5'-TGA GCC ACT GCA CCT G-3'		
5'-AGG CAT GAC TCA CCG C-3'		
5'-TTC TCA AGG TTC GTC CAT GT-3'		
5'-CCC AAC AGC AAT GGG AAG TT-3'		
5'-GAG GTG CCC GCT AGT A-3'		
5'-AGC TGA GAG CGC ATG TAT AA-3'		
5'-CAG AGA GCA AGA TCC TAC CTC-3'		
5'-TCC AGA GTC AAA AAC ACA GG-3'		
5'-CGT GAT TTC ATT TCT TGC TG-3'		
5'-TAG GCT TTG TTC TGG GGT TC-3'		
5'-GCA GGA AAT CGC AGG AAC TT-3'		
5'-GGC CCA GTT CAT TTT CTA GC-3'		
5'-TCT TTG ACC CAG ACC TCT AA-3'		

34/50

**FIG. 7I-1**

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D11S928	(277-289)	71%	11
D17S849	(251-261)	67%	17
D19S217	(219-233)	76%	19
D11S935	(196-208)	74%	11
D12S90	(166-182)	73%	12
D12S105	(137-155)	72%	12
D11S912	(101-123)	81%	11
<b>SET B</b>			
D21S261	(296-304)	50%	21
D19S220	(265-283)	84%	19
D12S88	(217-255)	85%	12
D7S480	(189-206)	80%	7
D15S120	(150-174)	74%	15
D10S210	(130-140)	80%	10
D20S119	(104-118)	82%	20
<b>SET C</b>			
D5S427	(280-302)	83%	5
D4S412	(237-249)	76%	4
D13S176	(211-227)	80%	13
D10S212	(189-201)	71%	10
D16S407	(150-170)	86%	16
D11S969	(141-149)	76%	11
D20S109	(106-133)	88%	20

35/50

# *FIG. 71-2*

## GROUP 9

---

### A Primer

---

5'-AAG TGA TCC ACC TGC CTT G-3'  
5'-CAA TTC TGT TCT AAG ATT ATT TTG G-3'  
5'-GGG GTT GAT TGA AGT TGG TT-3'  
5'-TAC TAA CCA AAA GAG TTG GGG-3'  
5'-AGC AGC AGC AGC CAT ATT GT-3'  
5'-TTT ACC TAA GGC TGG ATC TG-3'  
5'-TCG TGA GAN TAC TGC TTT GG-3'

5'-AAA ACA CCT TAC CTA AAA CAG CA-3'  
5'-ATG TTC AGA AAG GCC ATG TCA TTT G-3'  
5'-TGC ACC ACA GCA TAC CAG TA-3'  
5'-CTT GGG GAC TGA ACC ATC TT-3'  
5'-TTT GTG ATG GTC TTT TAT AGG CAT A-3'  
5'-CCT CAA TGC ACA ACT CCT-3'  
5'-CTG ACGA CAG TTT CAG TAT CTC TAT C-3'

5'-GCC TTC ACT AAG CAA TCT CTA AA-3'  
5'-ACT ACC GCC AGG CAC T-3'  
5'-CTG TGG GAT TCC TTA GTG ATA C-3'  
5'-GAA GTA AAG CAA GTT CTA TCC ACG-3'  
5'-CTC GCG CTG GGT ACA GTT AT-3'  
5'-TTG ATT TGG AAG ATT TTC AC-3'  
5'-AAC ACA CAT ACA AAC ACA CGC AGA T-3'



36/50

*FIG. 71-3*

B Primer	Annealing Temp.	Labeled Primer
5'-GCC TCT GAG AAT TAG TGT CTG TC-3'		
5'-CTC TGG CTG AGG AGG C-3'		
5'-CAA GAC CCA TAC CCA TGA-3'		
5'-CTA TCA TTC AGA AAA TGT TGG C-3'		
5'-AGT CAG GCC CAC CCA ATT TA-3'		
5'-CAA AGT TGA CAC TGA TTA TAG CA-3'		
5'-TTT TGT CTA GCC ATG ATT GC-3'		
5'-AGA TGA TGG TGA GTC CTG AG-3'		
5'-TCC CTA ACG GAT ACA CAG CAA CAC-3'		
5'-AAT GAA CAG CAA AAA CTA AGG GA-3'		
5'-AGC TAC CAT AGG GCT GGA GG-3'		
5'-GGC TCA AAG TGT TTG CAC TG-3'		
5'-CTC AGA CCT GGG TCA AGA TA-3'		
5'-TTT CCA GAT TTA GGG GTG TAT G-3'		
5'-ACA TGC TCT GAA TCA CCT GA-3'		
5'-CTA AGA TAT GAA AAC CTA AGG GA-3'		
5'-ATA TTC AGA CAA AAG CCA AGT TA-3'		
5'-TCT GTG TAC GTT GAA AAT CCC-3'		
5'-AGA TCA GAG GAG TGG GTT CC-3'		
5'-GGG GCA GAA TGG GTA T-3'		
5'-TTC CAG ACA GGA CAG CCT GC-3'		

SUBSTITUTE SHEET (RULE 26)

37/50

*FIG. 7J-1*

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D5S416	(282-292)	78%	5
D8S271	(257-271)	78%	8
D7S523	(224-240)	80%	7
D8S260	(187-213)	83%	8
D7S550	(177-200)	83%	7
D7S507	(148-168)	90%	7
D7S526	(125-135)	72%	7
D7S484	(99-113)	74%	7
<b>SET B</b>			
D20S106			20
D10S220	(267-291)	84%	10
D8S279	(229-257)	88%	8
D9S197	(199-215)	68%	9
D15S114	(177-187)	70%	15
D15S125	(157-169)	79%	15
D8S264	(121-145)	84%	8
<b>SET C</b>			
D8S263	(275-289)	75%	8
D9S166	(233-261)	82%	9
D13S164	(208-219)	72%	13
D9S164	(187-199)	80%	9q
D17S800	(168-178)	74%	17
D2S207	(144-156)	71%	2p
D9S161	(119-135)	78%	9

38/50

*FIG. 7J-3*

B Primer	Annealing Temp.	Labeled Primer
5'-AGT GAA ACT CGG NCC CTA-3'		
5'-AAC AAA CTT GCT TAT GAG TGT TAC T-3'		
5'-AAA ACA TTT CCA TTA CCA CTG-3'		
5'-GCT GAA GGC TGT TCT ATG GA-3'		
5'-GCA GTT GGG TTA TTT CAA GTC-3'		
5'-CTA CGT ACA TGG CTG CAA-3'		
5'-CCA TCT TGG TGT GAG GGC-3'		
5'-GCT GAG CAA GGC ATT GTT T-3'		
5'-ACT GAG GTC ATG CAA GAG GC-3'		
5'-GAG CAA GAC TGC ATC TCA AA-3'		
5'-GTG TCA GGT CGG GGT G-3'		
5'-ACG ATT TCT GGG AGA CTA TAT TGC-3'		
5'-TTG TCA CTG CTT TTC TCT GC-3'		
5'-CCC CTG AAG ACC GTG A-3'		
5'-CCA ACA CCT GAG TCA GCA TA-3'		
5'-ATG TAA CAA AAT GGA GTC GG-3'		
5'-TCC TAA TTC ACT GGG AAA AC-3'		
5'-ATT ACA GGC GTG ACA CAC C-3'		
5'-GTT TGC CTG GGG ATT GAT TT-3'		
5'-ATA GAC TGT GTA CTG GGC ATT GA-3'		
5'-ATG AAG AAA TAT ATA CAG TGC CG-3'		
5'-CAT GCC TAG ACT CCT GAT CC-3'		

39/50

*FIG. 7K-1*

Marker	Alleles (bp)	Heterozygosity	Chromosome
<b>SET A</b>			
D5S408	(247-299)	73%	5
D9S180	(220-265)	63%	9
D5S414	(186-206)	82%	5
D1S304	(168-206)	60%	1
D6S344	(139-159)	72%	6
D12S76	(112-124)	71%	12
D10S219	(89-103)	76%	10
<b>SET B</b>			
D11S906	(291-303)	73%	11
D15S121	(258-264)	66%	15
D5S425	(224-248)	77%	5
D5S395	(189-213)	81%	5
D13S217	(160-174)	67%	13
D2S206	(123-151)	79%	2
D6S263	(90-114)	81%	6
<b>SET C</b>			
D14S74	(291-313)	79%	14
D20S98	(259-275)	79%	20
D9S168	(227-247)	75%	9
D16S421	(206-212)	56%	16
D13S173	(166-178)	82%	13
D8S261	(128-148)	77%	8
D9S178	(93-99)	66%	9

40/50

**FIG. 7K-2**  
**GROUP 11**

---

---

**A Primer**

---

5'-ACA ACT TCC AAC CCT GAG AT-3'

9'-CAG TGG TTT GGA ATC GAA CC-3

5'-GGC CAG TTC AGT CAA GTG-3'

5'-ACC CTT TTT CCT CCA ATC AT-3'

5'-CTC CAG CCT GGG TCA CTA-3'

5'-GGG CTA CAT GAT GAG ACC CT-3'

5'-TCT TTC TAC CAC CCC CC-3'

5'-AGC TGG GCA CCG ATA GTA GT-3'

5'-TTG TAT CAG GGA TTT GGT TA-3'

5'-CTC CAG CCT GCT GAC C-3'

5'-GCA GAT GGA AAA CAC CAC TT-3'

5'-ATG CTG GGA TCA CAG GC-3'

5'-TTA AAA ATT AAG TAG GCT TTT GGT T-3'

5'-CTT AAG GCA AAA TTC TTT TCA ACA C-3'

5'-CCT GTA CCA CTA CCT GAG TTG AGT-3'

5'-GAA CTT GCA TAA CCC GAA T-3'

5'-GGT TTG TGG TCT TTG TAA GG-3'

5'-ACA TGA ACC GAT TGG ACT GA-3'

5'-CCC TGT TCC AGT AAT GAT GAC C-3'

5'-TGC CAC TGT CTT GAA AAT CC-3'

5'-GAA TAA AAC AGG GTT TGG G-3'

41/50

*FIG. 7K-3*

B Primer	Annealing Temp.	Labeled Primer
5'-ACT GTG CCT AGC CTT CAT TT-3'		
5'-AGC TAT TTT TGG GGG CTG AG-3'		
5'-TGG TTC CAG CAT ATA GCG-3'		
5'-AGA AGC TGA AAG CTG AGT GG-3'		
5'-CTA ATG CAT GAC AAT AAT ATT TCC A-3'		
5'-GCG GAG CTT CTT TTC TGT TG-3'		
5'-GCA GAG AAC CTA AAG CAT CC-3'		
5'-GCA CAG GCA AAG ANG AGG TA-3'		
5'-TGT TGT CGC TTC AGT ACA TA-3'		
5'-TCT TGG GCA AGC CAT C-3'		
5'-ACC TGC TGC TGG AAG ATT AC-3'		
5'-AAC CTG GTG GAC TTT TGC T-3'		
5'-GTC CTC ATG TGT TTA TGC TGT-3'		
5'-CTC AAA GTA AGA CCA TAA AAT ACC A-3'		
5'-CTT TGG CTG CCC GAA A-3'		
5'-CAA GGG TAT GTT CCC CAA AA-3'		
5'-TGG TTT GTT TGT ATA ACT ATC AT TG-3'		
5'-CCG TTCC CTA TAT TTC CTG G-3'		
5'-GTC TCT GGC TGC TCT CAA GAC TAT-3'		
5'-TAT GGC CCA GCA ATG TGT AT-3'		
5'-TTT CTC TAA GAA CTT TGG GG-3'		

42/50

*FIG. 7L-1*

---

Marker	Alleles (bp)	Heterozygosity	Chromosome
--------	--------------	----------------	------------

---

**SET A**

D19S209	(206-272)	77%	19
D14S77	(203-251)	92%	14
D10S189	(180-188)	72%	10
D12S87	(142-168)	79%	12

D13S158	(99-113)	81%	13
---------	----------	-----	----

**SET B**

D11S931	(251-267)	73%	11
D16S415	(208-234)	72%	16
D11S925	(173-199)	84%	11
D16S409	(135-147)	70%	16
D13S219	(117-127)	64%	13
D22S284	(86-102)	76%	22

**SET C**

D13S157	(250-264)	72%	13
D14S78	(211-233)	66%	14
D13S168	(173-197)	76%	13
D15S122	(143-159)	77%	15
D18S70	(111-126)	83%	18

43/50

## *FIG. 7L-2* GROUP 12

---

### A Primer

---

5'-TTC ATT CAC AAA TCN ATG GC-3'

5'-GCG TGA GTC ACT GTG CC-3'

5'-CAA AAG TAA CCA TTG AGC CC-3'

5'-CAC TAG GTG ATG CTG GAC AT-3'

5'-GTA CCC ACG GAG TGA AAG AA-3'

5'-GAT TGC TTG AGC CCA G-3'

5'-CCA GTA ATG TTA TGT AAG TCA ATG C-3'

5'-AGA ACC AAG GTC GTA AGT CCT G-3'

5'-TGA ATC TTA CAT CCC ATC CC-3'

5'-AAG CAA ATA TGC AAA ATT GC-3'

5'-ATG GGT ATT TAA CTT CTC TAC ACA G-3'

5'-AGC TGA GAA ATC ACA ACA GAG A-3'

5'-GGC AGG GAT AAG TAT GTC CT-3'

5'-GCC TAG CCC AGT GGT G-3'

5'-GAT AAT CAT GCC CCC CA-3'

5'-AAG GCT GAN CTC TAC CG-3'



44/50

*FIG. 7L-3*

B Primer	Annealing Temp.	Labeled Primer
----------	--------------------	-------------------

5'-CTG GAG AGC ATA GAC GNA GA-3'

5'-CAG ACA GAA ATT AAC CAG AGT TGA A3'

5'-TTG ATA GAA GAA GCG ATA GAT CG-3'

5'-CTG CAC AAA CAC TTG AAA CA-3'

5'-GCT TTG ACA ATT TAG CAG CA-3'

5'-GAG AAA TAG TAT GTG TTT GCC-3'

5'-TAG CCA CTG TAC CCC AGC-3'

5'-TTA GAC CAT TAT GGG GGC AA-3'

5'-AGT CAG TCT GTC CAG AGG TG-3'

5'-TCC TTC TGT TTC TTG ACT TAA CA-3'

5'-GCT CTC TTG AGG TCG TTA CA-3'

5'-TGG AAA TTT GCT GAC AGT AGA T-3'

5'-AAA GGT AAC ATC CAA GGG GT-3'

5'-TGC TTG TGC CTA TGT TCT TG-3'

5'-CCC AGT ATC TGG CAC GTA G-3'

5'-GGA ATG TCA AGA AGT ACC TAC CAT A-3'

45/50

**FIG. 7M-1**

---

Marker	Alleles (bp)	Heterozygosity	Chromosome
--------	--------------	----------------	------------

---

**SET A**

D13S156	(272-286)	80%	13
D19S226	(235-263)	84%	19
D16S422	(188-212)	78%	16
D18S65	(168-178)	71%	18
D16S413	(131-149)	83%	16

D20S95	(82-100)	83%	20
--------	----------	-----	----

**SET B**

D22S279	(249-258)	73%	22
D19S222	(233-241)	65%	19
D6S281	(203-219)	67%	6

D17S808	(147-167)	67%	17
---------	-----------	-----	----

**SET C**

D21S260	(267-277)	51%	21
D19S218	(240-256)	60%	19
D22S280	(208-220)	81%	22
D17S799	(186-200)	68%	17
D19S210	(165-177)	73%	19
D11S922	(88-138)	92%	11

SUBSTITUTE SHEET (RULE 26)

46/50

**FIG. 7M-2**  
**GROUP 13**

---

**A Primer**

---

5'-ATT AGC CCA GGT ATG GTG AC-3'

5'-CCA GCA GAT TTT GGT GTT GTC TA-3'

5'-CAG TGT AAC CTG GGG GC-3'

5'-GAG GCA GGA AAT TGC AGT GT-3'

5'-ACT CCA GCC CGA GTA A-3'

5'-AAA GCA AGG CTT CGT CTT AA-3'

5'-GCG ATC CAG CCT GTG T-3'

5'-GAA ATG TCC TAT TTG AAA CTG TGC-3'

5'-CTG GTA GTG TCA GGC ATG GC-3'

5'-ACC CTA GAC AGG ATG CCA-3'

5'-AGC TGT TCA TGC TTC CAT CT-3'

5'-TTT GCA TTT TCT GGA GTT TT-3'

5'-GCT CCA GCC TAT CAG GAT G-3'

5'-ATT GCC AGC CGT CAG TT-3'

5'-TCA CAC TCA CTG GTC TCT CA-3'

5'-GGG GCA TCT TTG GCT A-3'

SUBSTITUTE SHEET (RULE 26)

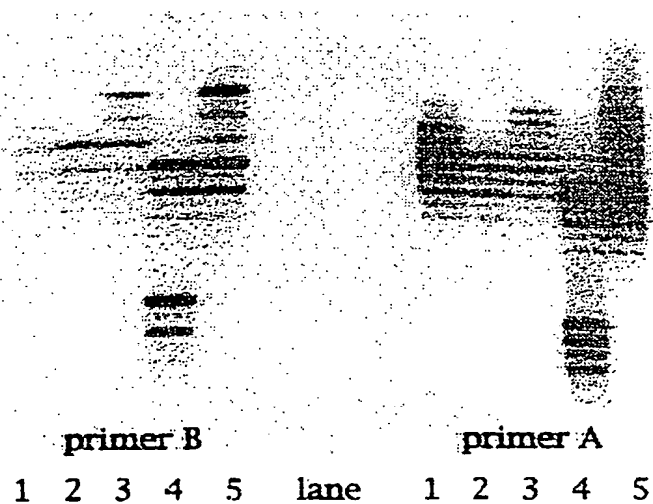
47/50

*FIG. 7M-3*

B Primer	Annealing Temp.	Labeled Primer
5'-GCT GTG GTA TGA GTT ACT TAA ACA C-3'		
5'-GGT CCA GGA TTT GAA CTA AAG CA-3'		
5'-CTT TCG ATT AGT TTA GCA GAA TGA G-3'		
5'-GCT GGT CTT ACT ATC TCA GGG G-3'		
5'-GGT CAC AGG TGG GTT C-3'		
5'-TTC NTC ATT TTA TTG TGT GCG-3'		
5'-TGT AAA TGG GGT AAG TGA TGC-3'		
5'-CTG TTG AAA TGT ATC CAG TAA ATC G-3'		
5'-CCT ATG TTT CAG GCA AAG GC-3'		
5'-TGT GGG TTT TCT CAG GTT AT-3'		
5'-AGA GCC CAG AAT ATT GAC CC-3'		
5'-AAT GTC CCT AAA CAC ATG GA-3'		
5'-GAT TCC AGA TCA CAA AAC TGG T-3'		
5'-GAC CAG CAT ATC ATT ATA GAC AAG C-3'		
5'-GGT GTG CCT GTG TGT AAA AG-3'		
5'-TCC GGT TTG GTT CAG G-3'		

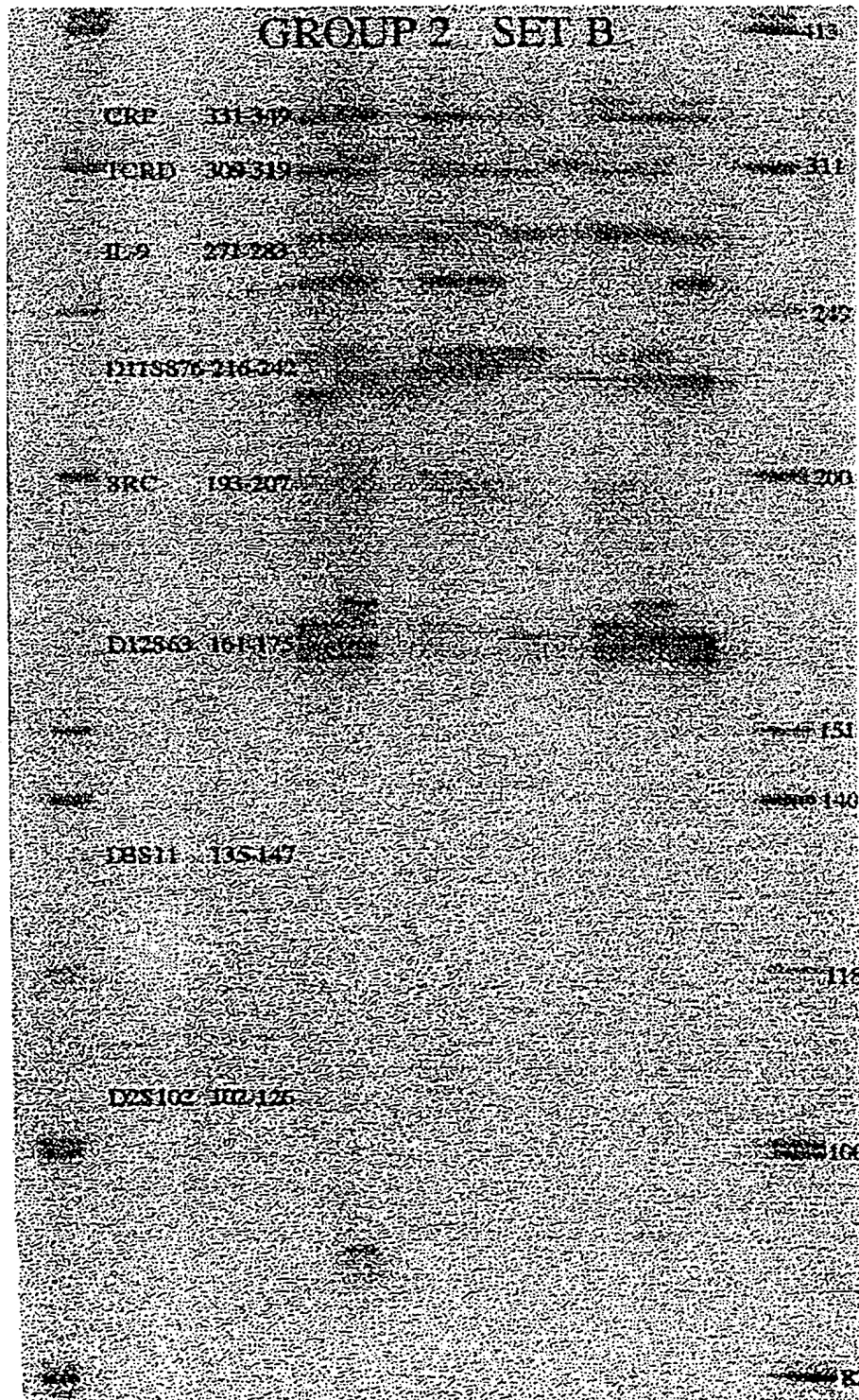
48/50

FIG. 8

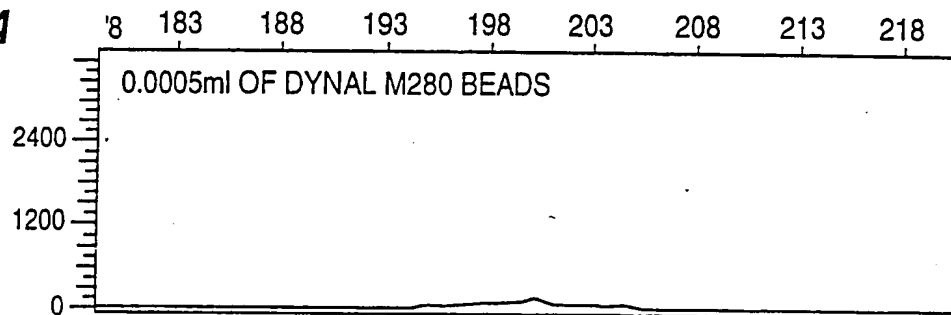
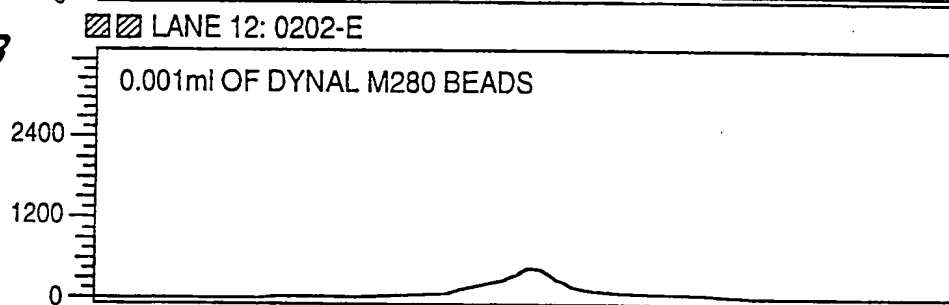
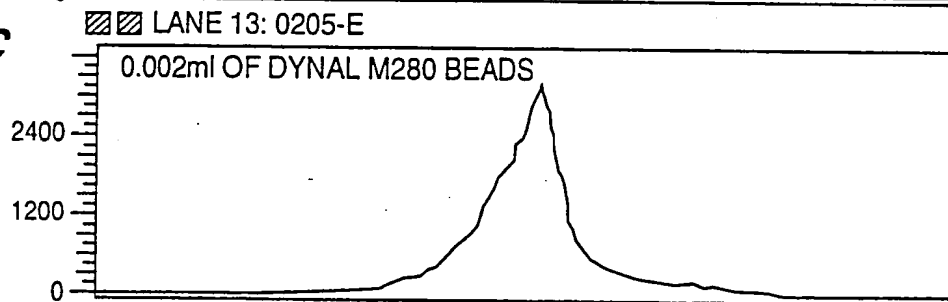
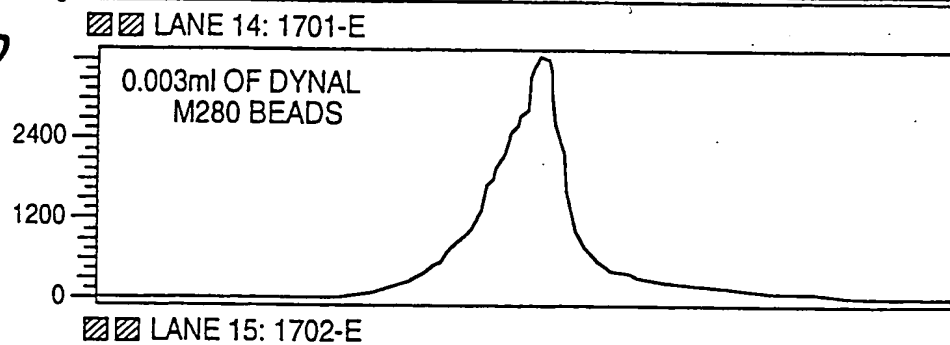


49/50

## FIG. 9



50/50

**FIG. 10A****FIG. 10B****FIG. 10C****FIG. 10D**

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/13945

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C12Q 1/68; C07H 21/02, 21/04

US CL : 435/6; 536 22.1, 24.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536 22.1, 24.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS, APS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	UK, A, 2 259 138 (JEFFREYS), 03 MARCH 1993, SEE ENTIRE DOCUMENT.	1-19
Y	US, A, 5,075,217 (WEBER) 24 DECEMBER 1991, SEE ENTIRE DOCUMENT.	1-19
Y	GENOMICS, VOLUME 14, ISSUED 1992, ZIEGLE ET AL, "APPLICATION OF AUTOMATED DNA SIZING TECHNOLOGY FOR GENOTYPING MICROSATELLITE LOCI", PAGES 1026-1031, SEE ENTIRE DOCUMENT.	1-19
Y	NATURE, VOLUME 359, ISSUED 29 OCTOBER 1992, WEISSENBACH ET AL, "A SECOND-GENERATION LINKAGE MAP OF THE HUMAN GENOME", PAGES 794-801, SEE ENTIRE DOCUMENT.	1-19

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 14 FEBRUARY 1995	Date of mailing of the international search report 03 MAR 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer <i>Deborah Freese for</i> EGGERTON CAMPBELL Telephone No. (703) 308-0196



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/13945

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GENOMICS, VOLUME 2, ISSUED 1988, SKOLNICK ET AL, "SIMULTANEOUS ANALYSIS OF MULTIPLE POLYMORPHIC LOCI USING AMPLIFIED SEQUENCE POLYMORPHISMS (ASPs)", PAGES 273-279, SEE ENTIRE DOCUMENT.	1-19
Y	AMERICAN JOURNAL OF HUMAN GENETICS, VOLUME 44, ISSUED 1989, LITT ET AL, "A HYPERVARIABLE MICROSATELLITE REVEALED BY IN VITRO AMPLIFICATION OF A DINUCLEOTIDE REPEAT WITHIN THE CARDIAC MUSCLE ACTIN GENE", PAGES 397-401, SEE ENTIRE DOCUMENT.	1-19
Y	AMERICAN JOURNAL OF HUMAN GENETICS, VOLUME 47, ISSUED 1990, DIEHL ET AL, "AUTOMATED GENOTYPING OF HUMAN DNA POLYMORPHISMS", PAGE A177, SEE ENTIRE DOCUMENT.	1-19